Looking back: Joke comprehension and the space structuring model

SEANA COULSON, THOMAS P. URBACH, and MARTA KUTAS

Abstract

We describe the space structuring model, a model of language comprehension inspired by ideas in cognitive linguistics, focusing on its capacity to explain the sorts of inferences needed to understand one-line jokes. One process posited in the model is frame-shifting, the semantic and pragmatic reanalysis in which elements of the existing message-level representation are mapped into a new frame retrieved from long-term memory. To test this model, we recorded participants' eye movements with a headband-mounted eye-tracker while they read sentences that ended either as a joke, or as nonfunny controls ("She read so much about the bad effects of smoking she decided to give up the reading/habit.") . Only jokes required frame-shifting; nonjoke endings were consistent with the contextually evoked frame. Though initial gaze durations were the same for jokes and nonjokes, total viewing duration was longer for the jokes and participants were more likely to make regressive (leftward) eye movements after reading the "punch word" of a joke. Results are consistent with the psychological reality of some process like frame-shifting, suggesting readers literally revisit aspects of the prior context while apprehending one-line jokes.

Keywords: Cognitive semantics; eye-tracking; frame-shifting; joke comprehension.

It often seems that life would be easier if people just said what they meant. For example, depending on the speaker 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

Humor 19–3 (2006), 229–250
DOI 10.1515/HUMOR.2006.013
0933–1719/06/0019–0229

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 and Ratcliff 1998). For instance, researchers in text processing have been concerned with the end-product of comprehension (Zwaan 1998), while those in sentence processing have focused more on how the message-level representation influences earlier processes. In short, mainstream research has primarily been concerned about how context influences the processing of individual words, rather than how words contribute to the message-level representation.

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 that 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. Because frame-shifting is a crucial aspect of joke comprehension, we review prior studies of joke processing, and present results of an eye-tracking experiment examining online joke comprehension.

1. Space structuring model

On the space structuring model (Coulson 2001), motivated by ideas in mental space theory (Fauconnier 1994), conceptual blending theory (Fauconnier and Turner 1998, 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 (such as Schank and Abelson 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 1983).

For example, if a listener heard the sentence in (1), she might, at one level at least, respond by constructing a model of the referential situation described by the speaker.

(1) When I asked the bartender for something cold and full of rum, he recommended his daiquiri.

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, among others. 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 might guide their expectations is perhaps best appreciated by examining examples in which those expectations are violated. For example, (2) is very similar to (1)—differing only in the final word—but prompts the construction of a very different cognitive model.

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

Rather than recommending a drink, the bartender in (2) 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 and new frame, the bartender's wife is accused of being a frigid lush. In fact, jokes are deliberately constructed to suggest one frame while evoking elements consistent with another.

Similarly, Raskin's (1985) Semantic Script Theory of Humor describes the narrative structure of jokes as involving the activation of two *scripts*, or frame-based event representations. Events in the joke are at least partially accounted for by both of the activated scripts, and jokes are funny when the two scripts are opposed to one another in particular ways, such as possible versus impossible, or real versus unreal. One shortcoming of this model is its basis in traditional assumptions about knowledge representation.

Scripts have been criticized for being overly rigid data structures that are ill-suited for the representation of unusual events (see, e.g. Allen 1987). 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 script itself (Wilensky 1986). In short, though knowledge of typical scenarios represented in scripts and frames is necessary for the comprehension of many jokes, it is far from sufficient. Thus the space structuring model appeals to processes proposed in cognitive semantics for the creative combination of frames and the construction of novel frames in response to contextual demands.

The implications of the space structuring model for joke comprehension can be seen as an extension of the general approach to meaning construction in cognitive linguistics. For example, in arguing that language is not completely compositional, Langacker (2000) notes that the meaning of a complex expression is often more specific than the meaning of its individual components, and can even conflict with those constituent meanings. Explaining this phenomenon, Langacker (2000: 15) writes:

> When a novel expression is first used, it is understood with reference to the entire supporting context. The speaker relies on this context, being able to code explicitly

only limited, even fragmentary portions of the conception he wishes to evoke. Usually, then, the expression's conventionally determined import at best approximates its actual contextual understanding ... it does not contain or convey the intended meaning, but merely furnishes the addressee with a basis for creating it (...).

Similarly, Fauconnier (1997) and Fauconnier and Turner (2002) explain how the very same expression can take on a vast number of remarkably different meanings in context. Given linguistic input in a particular context of use, speakers exploit cognitive operations such as metaphorical mapping and conceptual blending to construct enriched meanings in context. Though recent modifications of Raskin's script-based model of joke comprehension represent improvements to the original approach (see, e.g. Attardo and Raskin 1991; Ruch et al. 1993), the processes and knowledge structures that underlie these models cannot capture the creative aspects of everyday meaning construction (Fauconnier 1997).

Although jokes highlight the human capacity for creative meaning construction, they are certainly not its only arena. The frame-shifting which is such a prominent aspect of narrative jokes can also be prompted by other situations. For example, in *The Sixth Sense*, Bruce Willis plays a child psychologist treating a disturbed boy who "sees dead people." At a critical moment in the film, it becomes apparent that Willis, himself, is dead, prompting the viewer to reevaluate Willis' relationship with the boy, as well as his seemingly estranged interactions with his wife.

Besides horror movies, frame-shifting is employed in political arguments. Coulson (2001) describes an argument in which an abortion rights advocate asks his opponent, "If one of your daughters is raped, should she be able to have an abortion?" The anti-abortion opponent replied, "I become a grandfather, then." The initial inquiry presents the man with a state of affairs concerning his daughter and invites him to respond to how he would feel about his daughter getting an abortion in these circumstances. In doing so, he evokes cultural models of rape and unjust punishment that are salient in the American debate on this topic. Though the reply may seem somewhat bizarre, it is actually a deft rhetorical move that requires frame-shifting in order to construe the statement's relevance to the on-going discourse. Regardless of whether the listener agrees with the man's construal of this scenario, to even understand his reply, it is necessary to rearrange the contextually established information by shifting to a kinship frame.

Even more commonplace examples of frame-shifting might include a case where the *wolf*, or an intruder in the house turns out to be the family dog. But 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 (3) the reader begins by evoking a frame in which 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.

The cue to shift frames is a single word (*years*), known as the *disjunctor* (Attardo et al. 1994). Upon encountering this cue at sentence end, the reader is forced to go back, at least mentally, and reinterpret the word *time* that appeared earlier in the sentence 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 so that the initial busy professional interpretation can be mapped into the crooked-businessman frame.

Frame-shifting involves a marked reorganization of the message-level aspects of the utterance, most of which cannot be attributed to compositional mechanisms of reanalysis. For instance, most examples of frame-shifts in jokes do not require the listener to instantiate a new structural analysis of the sentence. Though the listener is led down a pragmatic garden path, pragmatic reanalysis often proceeds without syntactic reanalysis. In (2), for example, *wife* is the object of *recommended* just as *daiquiri* is in (1), the straight version of (2). In (3), 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, novel meanings particular to a specific context. For example, in the accountant joke in (3) "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 (3) as in "do time" is only congruous because of a stereotyped scenario that involves accountants obscuring illegal business dealings.

2. The psychological reality of frame-shifting

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 contributions can easily be accommodated by the activated frame will be easier to process than words unrelated to it. 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. The relationship between a word and its context 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 that differs from that presented by lexical violations which are consistent with the currently active frame.

2.1. *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, individuals 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 can record how long participants 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 disjunctor, 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 (Taylor 1957).

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 were more constraining than others. For example, (4) elicited a similar response from 81% of the participants, while (5) elicited many different responses, albeit mostly from the gambling frame.

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

(5) 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 (4) which elicited at least one response with a cloze probability of greater than 40%, and low constraint jokes like (5) 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 (4) was *name* (the joke ending was *dress*); while the straight ending for (5) 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%. Moreover, joke and non-joke endings were closely matched for factors such as word length and word frequency that are known to affect how long it takes to read a word.

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 did indeed spend 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 may be 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 ongoing sentence.

While psycholinguists have traditionally approached meaning construction from the perspective of how low-level processing of words is influenced by the developing message-level representation, this study suggests we also need to consider how lexical processing in turn may influence the development of the message-level representation. In Coulson and Kutas' (1998) study, the jokes and nonjoke controls differed only in their last word. Longer reading times for the jokes, then, must somehow reflect differences in the complexity of high-level processes of meaning construction. The apparent processing cost attributed to frame-shifting is consistent with our suggestion that lexical processing triggers the creative construction of cognitive models in working memory.

2.2. *Event-related brain potentials*

Another way of assessing readers' on-line comprehension of language materials is to use event-related brain potentials (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 simultaneously, they create an electrical field in the brain that can be detected at the scalp via the electroencephalogram (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.

Over thirty 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 et al. (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 1997).

In a classic ERP language experiment, Kutas and Hillyard (1980) contrasted ERPs elicited by visually presented sentences that ended congruously, as in (6), with ERPs elicited by sentences that ended incongruously, as in (7).

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

(7) 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.

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 (Kutas et al. 2000). 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 (6), large for low cloze completions like *socks* in (7), 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 study discussed above. These two kinds of studies 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 and Kutas 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 (as in Coulson and Kutas 1998). Jokes elicited larger amplitude N400 than nonjoke controls — but only in the high constraint sentences. Joke and nonjoke endings were both unexpected words. However, because nonjoke endings were designed to be congruent with the contextually evoked frame, the ending type manipulation affected consistency with discourse-level expectancies. The observed differences in ERPs elicited by jokes and nonjoke controls thus demonstrate the brain's sensitivity to high level expectations such as those based on frames. In low constraint sentences, the general nature of the scenario may be somewhat unclear until the last word. As a result, the nonjoke endings to these low constraint sentences are less likely to be facilitated, either because readers do not completely commit to a particular frame, or because they commit to a different frame than the one we, as experimenters, chose. Our finding of joke effects on the N400 component seems to indicate that joke comprehension poses additional demands on neural processes generally associated with the processing of meaningful events.

Joke endings also elicited larger positivities in the ERPs 500–900 ms after the word was presented (after the N400 component). Relative to controls, high constraint jokes elicited a positivity evident over posterior regions that resembled an ERP effect (the P3b) associated with surprising events (see also Derks et al. 1997). Low constraint jokes elicited a positivity over frontal regions that resembled an ERP effect (the P3a) associated with novel events that evoke an orienting response. Coulson and Kutas (2001) suggested the posterior positivity reflects the violation of frame-level expectations set up by the high constraint sentence contexts, while the frontal positivity may reflect the perception of a stimulus from a completely novel category set up by low constraint sentence contexts.

The ERP effect whose behavior is most consistent with the frame-shifting process was a sustained negativity 500–900 ms focused over left lateral frontal electrode sites. This effect was observed only in participants who consistently got the jokes. Further, among these good comprehenders, the sustained negativity was evident in both high and low constraint stimuli, perhaps indexing the additional processing required for joke comprehension. Although spatial localization of ERP effects is inherently difficult, the markedly focal nature of the sustained effect is at least consistent with a generator in dorsolateral prefrontal cortex, and may index prefrontal activation implicated in the operation of verbal working memory. Coulson and Kutas (2001) suggest that the sustained negativity indexes the rebinding of discourse elements in working memory.

Because the sustained negativity 500–900 ms was elicited by both high and low constraint jokes in good, but not poor joke comprehenders, it was argued to index frame-shifting needed to establish a coherent interpretation of the joke. Interestingly, although the morphology (wave shape) and scalp topography of this sustained negativity clearly differentiate it from the positivities in the ERP to high and low constraint jokes, the three sorts of effects do occur within the same time window. Temporal overlap of the ERP indices of surprise- and coherence-related processing suggest that these two stages may not be as distinct from one another as has been assumed in traditional accounts of joke processing (Suls 1983).

Consistent with this latter observation, Vaid and colleagues have found that during the intermediary phase of jokes, words related to both the joke and the non-joke interpretations of the scenario are primed, though afterwards only the joke meaning is active (Vaid et al. 2003). For example, in the joke "I tried to sniff Coke, but the ice cubes got stuck in my nose," while participants' response to both *drug* and *soda* was facilitated after "ice cubes," only *soda* was primed after the offset of the joke. These findings suggest that the registration of surprise and the search for an alternative interpretation are not discrete sequential processing events, but rather parallel ones with overlapping time courses. The concurrent activation of joke- and non-joke-related meanings during the incongruity phase of jokes may reflect a general strategy whereby the language processor searches for salient alternative interpretations.

3. The present study: Eye movement registration

The self-paced reading paradigm is a good technique for establishing that one kind of sentence takes more time to read than another (presumably very comparable) type of sentence. This technique, however, lacks 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 drawback 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 might have engendered more processing difficulty. However, as noted above, 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, we 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 (as in Coulson and Kutas 1998, 2001).

3.1. *Methods*

3.1.1. *Participants.* Participants were 32 undergraduate students from the University of California, San Diego, who received course credit for their participation.

3.1.2. *Materials and design.* Experimental materials were identical to those used in Coulson and Kutas (2001), consisting of 60 pairs of sentences (jokes and their controls). Half of these stimuli (30 sentences) were deemed high constraint, half low constraint. Two lists of stimuli were constructed so that while no individual participant saw both versions of the same sentence, across participants all stimuli occurred equally often as jokes and as non-funny controls. Each participant read 30 jokes (15 high constraint, 15 low constraint) and 30 nonjoke controls (15 high constraint, 15 low constraint). To encourage deep processing of the stimuli, all sentences were followed by a yes/no comprehension question.

3.1.3. *Apparatus.* Movements of the left eye were recorded using the EyeLink headband-mounted eye-tracking system. Both X and Y coordinates were collected at a sampling rate of 250 Hz and a spatial resolution of <1 degree of visual angle.

3.1.4. *Procedure.* Participants were tested individually in a session lasting under two hours. The session began with calibration of the eye tracker, and the calibration procedure was repeated after each block of sentences. Participants' task was to read the sentences, and answer the yes/no comprehension question that followed each sentence.

3.1.5. *Results and discussion.* Experimental sentences were divided into three regions of one or more words for analysis. The first two regions of interest were chosen based on results of a separate normative study in which 16 people from the same population as our participant pool were given a printed list of jokes from the study and asked to underline information in the sentence that was critical for getting each joke. For each joke we chose the two portions of the sentence that participants most consistently underlined. The first region occurred early in the sentence, while the second region typically included words leading up to the sentence-final word. Because all jokes turned on the sentence-final word, participants almost always underlined it. However, to better compare our results with those of the self-paced reading time studies, we considered the sentence-final word in each sentence to comprise the third region of interest. In the examples below, the regions of interest have been italicized.

She went on a *fourteen day diet* [R1] but she only *lost two* [R2] *weeks./ounces*. [R3]

I knew a *streetwalker who was so exclusive* [R1] she had an *unlisted telephone* [R2] *booth./line*. [R3]

She *read so much about the bad effects of smoking* [R1] she decided to *give up the* [R2] *reading./habit*. [R3]

For each region, two measurements were made: *gaze duration*, defined as the amount of time spent reading the word the first time through the sentence, *total viewing duration*, or the sum of the gaze duration and the time that elapsed during any subsequent fixations in the region. Because we were particularly interested in whether jokes would differentially provoke regressive eye movements, for Region 3 we also measured *regression probability*, the probability of making a leftward eye movement after one's initial fixation of the word or words in that region. To determine whether they were statistically significant, results were analyzed with repeated measures analysis of variance (ANOVA) with factors Sentence Type (Jokes/Controls), and Sentence Constraint (High/Low).

In Region 1, there were no effects of Sentence Type on either gaze duration or total viewing duration, as participants spent approximately the same amount of time reading words in the jokes as in the control sentences. There was, however, a main effect of Sentence Constraint on both gaze duration ($F(1,31) = 11.8$, $p < 0.01$) and total viewing duration ($F(1,31) = 28.8$, $p < 0.001$). Given that different words occur in the high and the low constraint sentences, respectively, it is not surprising that participants took longer to read words in one group of sentences than the other (in this case, the high constraint sentences).

In Region 2, there were no effects of either Sentence Type or Sentence Constraint on gaze duration, indicating that participants spent the same amount of time reading words in jokes and controls the first time they passed through this region. Again this is unsurprising, given that participants cannot know whether or not they are reading a joke until they reach the sentence-final word (Region 3). For this reason it is significant that both Sentence Type ($F(1,31) = 5.7$, $p < 0.05$) and Sentence Constraint ($F(1,31) = 7.1$, $p < 0.05$) reliably affected total viewing durations. The total viewing duration was longer for words in the high than the low constraint sentences, and was longer for words in jokes than controls (see Figure 1).

In Region 3, as in Region 2, there was no effect of either Sentence Type or Sentence Constraint on participants' initial gaze duration on sentence-final words. There was, however, a trend towards longer total viewing durations for jokes (560 ms, $se = 34$ ms) than for the non-funny control stimuli (498 ms, $se = 25$ ms), and for the last word of low (558 ms, $se = 29$ ms) than high (500 ms, $se = 31$ ms) constraint sentences. However, neither the Sentence Type ($F(1,31) = 2.8$, $p = .10$) nor the Sentence Constraint ($F(1,31) = 3.8$, $p = .06$) effect was reliable. Finally, though

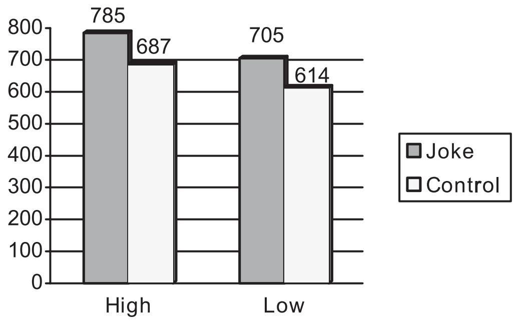

Figure 1. *Region 2: Total viewing duration in ms*

regression probability did not differ as a function of Sentence Constraint, participants were significantly more likely to make a regressive eye movement after a joke (57%, se = 3.5%) than a non-funny ending (52%, se = 3.5%) to the same sentence ($F(1,31) = 4.7$, $p < 0.05$).

Thus while participants initially spent the same amount of time reading Region 2 and Region 3 in jokes and the controls, they were more likely to make regressive eye movements after the last word of jokes. Although this difference was quite small (only 5%), it was quite robust across subjects. Moreover, when they did so, they tended to return to (or re-fixate) the words in Region 2 and, (though less reliably), the sentence-final word in Region 3 when the sentence ended as a joke. Thus while initial gaze durations were the same for jokes and controls in Regions 2 and 3, total viewing duration tended to be longer for the jokes because of the greater probability and length of subsequent fixations in these regions.

Interestingly, participants spent longer reading words in Regions 1 and 2 in *high* than low constraint sentences, but longer reading words in *low* constraint sentences in Region 3. This may be because in high constraint sentences the important information occurs early, while in low constraint sentences the last word provides crucial content. Presumably the reason people exhibit a high degree of consensus about the last word of a high constraint sentence is because information presented early in the first part of the sentence significantly winnows the set of possible endings. So, even though the sentence-final words in high and low constraint sentences were equally unexpected (that is, they always tended to be words that few, if any, people produced on the cloze task), people spent longer reading the final word (Region 3) of low constraint sentences because they were more likely to provide new message-level information.

One difference between these results and those in the self-paced reading (Coulson and Kutas 1998) and ERP (Coulson and Kutas 2001) studies that used the same materials was that we failed to replicate the interaction between Sentence Ending and Sentence Constraint. That is, in our previous studies, differences in joke and nonjoke endings were larger and more robust for high than low constraint sentences. Given that both self-paced reading and the word by word presentation used in the ERP study involved somewhat unnatural reading conditions, one possibility is that the earlier results are artifactual in nature. Another possibility is that our total viewing durations for the last word (Region 3) were too variable across participants to reach statistical significance. Consistent with this suggestion, the pattern observed in the present study was similar to that observed in our previous self-paced reading study: in high constraint sentences, participants spent 103 ms longer reading the joke endings than the controls; in low constraint sentences, this difference was only 21 ms.

In the present study, we found that just as in the reading time study, readers spent reliably longer reading sentences that ended as jokes. Moreover, this difference in reading times was not observed for gaze durations, but rather total viewing durations. These findings argue against the possibility that joke effects result simply because context facilitates the low-level processing of nonjoke endings. Rather, joke effects arose in the later stages of processing associated with subsequent fixations of the sentence final word. Further, in both high and low constraint sentences, participants were more likely to look back to earlier words in the sentence when they encountered a joke ending than a straight one. This finding is consistent with the psychological reality of frame-shifting, suggesting readers literally revisit aspects of the preceding context in order to activate a new frame so as to better get the jokes. However, given the small size of the observed effects, one might question whether the processes involved in joke comprehension differ in kind or in degree from those involved in understanding our controls. Future research should involve more fine-grained analysis to determine whether there are systematic differences in which region or series of regions are re-examined in jokes and nonjoke controls, respectively.

4. Summary

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 and Kutas 1998). Similarly, ERP results from the study by Coulson and Kutas (2001) also suggest the differential, presumably costly, processing 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 is consistent with our finding in the self-paced reading study that joke effects 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 ongoing construction of the message-level representation.

The present study confirmed that the processing cost of frame-shifting was evident under more natural reading conditions. Moreover, the eye movement data suggested that the cost was not at the level of word recognition (indexed by the length of a reader's initial fixation of a word), but rather 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). Importantly, we found that people were more likely to make regressive eye movements when they read the joke than the straight endings, as if they wanted to reexamine earlier parts of the sentence for clues to which alternative frames should be retrieved.

These experiments suggest that the relationship between a word and its surrounding context is multifold. This relationship involves the ways that individual words add to the cognitive models active in working memory and how individual words can prompt the construction of new models. The space structuring model correctly 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. 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 even with minimal linguistic input.

Of course, the finding that discourse-level considerations affect the processing of individual words is counterintuitive only from the traditional building blocks approach to meaning construction. That is, once we abandon the notion that contextual and background knowledge are brought to bear after the assembly of a context-invariant meaning, the finding that the same factors operate at the lexical, sentential, and intersentential levels should come as no surprise. If language is designed to prompt the construction of cognitive models, the cuing of projections, and so forth, we should actually predict that words, sentences, and groups of sentences also can and routinely do prompt the same sorts of operations.

Because language use, in particular, is firmly rooted in human experience and social interaction, we need to construe meaning construction as a set of routines for assembling cognitive models that enable interpretation, action, and interaction. Besides acknowledging the crucial role of the physical and social world within which we function, the space structuring model is congruent with rising consciousness in cognitive neuroscience of the importance of the motor system, and the growing realization that attention, perception, and memory are all intimately connected with action.

Neuroscience gives us a picture of information processing as involving partitioning of sensory information into parallel streams, each computing different sorts of information, and each with its own hierarchical structure. The massively interconnected systems allow for information to be continuously mapped and remapped between intertwined processing streams. Similarly, the space structuring model, though motivated by very different issues and sorts of data, portrays meaning construction in an analogous way: the partitioning of information into parallel streams, extensive mapping, and the integration of disparate information needed for adequate message-level comprehension. While the establishment of abstract mappings in mental space theory is not directly comparable to mapping in the visual system, perhaps computationally similar mechanisms of information regulation underlie the flexibility evident in both meaning construction and visual processing.

In fact, recent research in cognitive psychology points to the import of what Barsalou (1999) calls perceptual symbols. Perceptual symbols are mental representations that are neither perceptual—that is, strictly dependent on sensory input systems — nor symbolic — that is, completely amodal. As outlined in Barsalou (1999), schematic representations of perceptual experience are stored around a common frame that promotes schematized simulations. Importantly, such simulations need not be accompanied by the experience of visual imagery, and are not to be construed as mental "pictures." Indeed, perceptual symbols recruit neural machinery activated in perceptual experience from all modalities — auditory, olfactory, somatosensory, and kinesthetic, as well as visual. As abstract perceptual experience, perceptual symbols develop to support categorization, inference, and interaction with the world around us. Frames built from perceptual symbols present themselves as representations that can sustain the creative processes of meaning construction necessary to understand jokes.

University of California, San Diego

Note

Correspondence address: coulson@cogsci.ucsd.edu

References

Allen, James
1987 *Natural Language Processing*. Philadelphia, PA: Benjamin/Cummings Publication.

Attardo, Salvatore, Donalee H. Attardo, Paul Baltes, and Marnie J. Petray
1994 The linear organization of jokes: Analysis of two thousand texts. *Humor: International Journal of Humor Research* 7 (1), 27–54.

Attardo, Salvatore, and Victor Raskin
1991 Script theory revis(it)ed: Joke similarity and joke representation model. *Humor: International Journal of Humor Research* 4 (3/4), 293–347.

Barsalou, Lawrence W.
1999 Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–609.

Coulson, Seana
2001 *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge, UK: Cambridge University Press.

Coulson, Seana, and Marta Kutas
1998 *Frame-Shifting and Sentential Integration* (Cognitive Science Technical Report No. 98.02). La Jolla, CA: UCSD.
2001 Getting it: Human event-related brain response in good and poor comprehenders. *Neuroscience Letters* 316, 71–74.

Derks, Peter, Lynn S. Gillikan, Debbie S. Bartolome-Rull, and Edward H. Bogart
1997 Laughter and electroencephalographic activity. *Humor: International Journal of Humor Research* 10 (3), 285–300.
Fauconnier, Gilles
1994 *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: Cambridge University Press.
1997 *Mappings in Thought and Language*. Cambridge, UK: Cambridge University Press.
Fauconnier, Gilles, and Mark Turner
1998 Conceptual integration networks. *Cognitive Science* 22, 133–187.
2002 *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
King, Jonathan, and Marta Kutas
1995 Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7 (3), 376–395.
Kutas, Marta
1997 Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology* 34 (4), 383–398.
Kutas, Marta, Kara D. Federmeier, Seana Coulson, Jonathan King, and Thomas F. Munte
2000 Language. In Cacioppo, John T., Louis G. Tassinary, and G. G. Berntson (eds.), *Handbook of Psychophysiology*, 2nd ed. Cambridge: Cambridge University Press, 576–601.
Kutas, Marta, and Steven A. Hillyard
1980 Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207, 203–205.
Langacker, Ronald W.
1987 *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
2000 *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter.
Mangun, George R., Steven A. Hillyard, and Steven J. Luck
1993 Electrocortical substrates of visual selective attention. In Meyer, David E. and Sylvan Kornblum (eds.), *Attention and Performance 14: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*. Cambridge, MA: MIT Press, 219–243.
McKoon, Gail, and Roger Ratcliff
1998 Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology* 49, 25–42.
Norman, Donald A.
1983 Some observations on mental models. In Gentner, Dedre and Albert Stevens (eds.), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum, 7–14.
Raskin, Victor
1985 *Semantic Mechanisms of Humor*. Dordrecht: D. Reidel.
Ruch, Willibald, Salvatore Attardo, and Victor Raskin
1993 Toward an empirical verification of the General Theory of Verbal Humor. *Humor: International Journal of Humor Research* 6 (2), 1233–1136.
Schank, Roger C., and Robert P. Abelson
1977 *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum.

Suls, Jerry
1983 Cognitive processes in humor appreciation. In McGhee, Paul and Jeffrey Goldstein (eds.), *Handbook of Humor Research*. New York: Springer-Verlag, 39–55.

Taylor, Wilson
1957 "Cloze" readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology* 41, 19–26.

Vaid, Jyotsna, Rachel Hull, Roberto Heredia, David Gerkens, and Francisco Martinez
2003 Getting a joke: The time course of meaning activation in verbal humor. *Journal of Pragmatics* 35, 1431–1449.

Wilensky, Robert
1986 Points: A theory of the structure of stories in memory. In Grosz, Barbara J., Karen S. Jones, and Bonnie L. Webber (eds.), *Readings in Natural Language Processing*. Los Altos, CA: Morgan Kaufman Publishing, 459–473.

Zwaan, Rolf A.
1998 Situation models in language comprehension and memory. *Psychological Bulletin* 123, 162–185.