Conceptual integration and metaphor:
An event-related potential study

SEANA COULSON
University of California, San Diego, La Jolla, California

and

CYMA VAN PETTEN
University of Arizona, Tucson, Arizona

Event-related brain potentials (ERPs) were recorded from 18 normal adults as they read sentences that ended with words used literally, metaphorically, or in an intermediate *literal mapping* condition. In the latter condition, the literal sense of the word was used in a way that prompted readers to map conceptual structure from a different domain. ERPs measured from 300 to 500 msec after the onset of the sentence-final words differed as a function of metaphoricity: Literal endings elicited the smallest N400, metaphors the largest N400, whereas literal mappings elicited an N400 of intermediate amplitude. Metaphoric endings also elicited a larger posterior positivity than did either literal or literal mapping words. Consistent with conceptual blending theory, the results suggest that the demands of conceptual integration affect the difficulty of both literal and metaphorical language.

Studied for centuries by rhetoricians, metaphor is considered the paradigmatic example of the *trope*—that is, a word used in its figurative sense (Aristotle, trans. 1952; Quintillian, trans. 1921). Itself somewhat metaphoric, trope is the Greek word for *twist*, or *turn*. Nonliteral language has traditionally been viewed as a deviation from normal language use and one that takes extra effort to understand. The *standard pragmatic model* (Grice, 1975; Searle, 1979) stipulates that (1) metaphors are "special" and consequently are processed with qualitatively different mechanisms than those for literal language, and (2) the computation of literal meaning precedes that of metaphoric meaning.

Cognitive linguists have attacked the *specialness* assumption by noting that metaphor is pervasive in everyday language and that it plays a pivotal role in historical language change (Lakoff & Johnson, 1980; Sweetser, 1990; Turner, 1991). Given systematic relationships between literal and metaphoric uses of the same words, Lakoff (1993) has suggested that metaphors reflect the output of a cognitive process by which we understand a *target* domain by exploiting cognitive models from an analogically related *source* domain. In *conceptual metaphor theory*, clusters of related expressions (e.g., *fuming, boiling, blow-*

ing one's top) are the manifestation of underlying conceptual metaphors (e.g., *anger = fluid in a heated container*). Lakoff has further argued that "the system of conventional conceptual metaphor is mostly unconscious, automatic, and is used with no noticeable effort, just like our linguistic system and the rest of our conceptual system" (pp. 227–228).

A variety of reaction time measures have indicated that metaphor interpretation is neither slow nor optional, casting doubt on the second tenet of the standard model. When the metaphoric interpretation of a sentence has adequate contextual support, metaphors are read no more slowly than literal language (Gibbs, Bogdanovich, Sykes, & Barr, 1997; Inhoff, Lima, & Carroll, 1984; Ortony, Schallert, Reynolds, & Antos, 1978). Futhermore, readers take longer to reject statements that are literally false but metaphorically true than to reject nonmetaphoric false statements (Gildea & Glucksberg, 1983; Glucksberg, Gildea, & Bookin, 1982; Keysar, 1989). This finding suggests that literal and metaphoric meanings become available simultaneously, thus producing response competition. Also, Blasko and Connine (1993) showed that following metaphors rated as *apt* (viz. readily interpretable), lexical decisions for target words related to figurative meanings were made just as fast as those for targets related to literal meanings. For example, after a phrase like *hard work is a ladder*, *advance* and *rungs* both received faster responses than did *pastry*. Because the target words were presented immediately after the offset of the last word of a spoken metaphor, these authors concluded that the figurative meaning was rapidly available.

In contrast to the standard model, current processing models of metaphor comprehension all assume that literal

and nonliteral language comprehension invoke the same mechanisms (Gibbs, 1994; Glucksberg, McGlone, & Manfredi, 1997; Wolff & Gentner, 2000). These mechanisms include one's noting the potential correspondence between semantic attributes or relational structure associated with the source and target domains (*alignment)* and a selective projection of properties from one to the other (Shen, 1999). Most models also assume that metaphor comprehension involves the selection of some attributes at the expense of others, a process previously described as necessary for the interpretation of both ambiguous and unambiguous literal words in context (Tabossi, 1991). Similarly, Gernsbacher and Robertson (1999) have suggested that metaphor comprehension necessitates suppression of irrelevant semantic attributes, but that the same general mechanism is invoked during the interpretation of anaphors, lexical ambiguities, and syntactically ambiguous phrases. In contrast to Lakoff's (1993) claim that metaphor processing is effortless, current processing models suggest that, ceteris paribus, metaphoric language places heavier demands on the mechanisms of alignment, selective projection, and inference than does literal language. For instance, Blasko (1999) writes, "If metaphor involves creating a bridge between dissimilar semantic domains and filtering out or suppressing unimportant characteristics while selecting relevant ones, then it should require considerable working memory capacity for both access and mapping processes" (p. 1679).

Surprisingly, data supporting the prediction that comprehension of metaphoric language should involve some extra effort is largely absent from psycholinguistic research. As is noted above, most studies suggest that when metaphors are preceded by sufficient context to be interpretable, literal and metaphoric language are processed in the same amount of time. However, equivalent processing times need not imply equivalent effort. By analogy, it may take the same amount of time to lift a 5- and a 20-pound weight, but the latter recruits more resources. The failure to demonstrate longer processing times for metaphoric language might also reflect a mismatch between the power of the dependent measures and the subtlety of the processing differences between literal and nonliteral language. In many studies, reading times for entire sentences or large sentence fragments have been found, so minor slowing on critical words might have gone undetected (e.g., Gibbs, 1990; Gibbs et al., 1997; Glucksberg et al., 1997; see a similar critique by Blasko, 1999). Frisson and Pickering (2001) have noted that word frequency, plausibility, and cloze probability have not always been adequately controlled in studies in which reading times for literal and figurative language are compared.

We suggest that the *continuity claim* (that literal and nonliteral language processing occur in the same time course and involve the same processing mechanisms) common to modern accounts of metaphor processing, is very different from the *equivalence claim* (that metaphoric language is no more difficult to comprehend than literal language). If the same operations are involved in literal and nonliteral lan-

guage comprehension (the continuity claim), principles governing the difficulty of metaphor comprehension ought also to apply to literal language. The goals of the present study were (1) to determine whether metaphors are more difficult to understand than literal sentences by investigating processing difficulty independent of reaction time, and (2) to evaluate the continuity of literal and metaphoric language by including a condition hypothesized to be midway between the overtly metaphorical and the clearly literal. Dubbed *literal mappings*, these instances of literal language impose similar, but lighter, demands on processes of mapping, selective projection, and conceptual integration as metaphor. Below, we explain the construction of this intermediate condition and explain why we used electrical brain activity as a dependent measure.

Conceptual Blending and Literal Mappings

Our selection of a condition midway between metaphoric and literal language was inspired by a general theory of conceptual integration known as *conceptual blending* (Fauconnier & Turner, 1998). As it pertains to metaphor interpretation, conceptual blending theory suggests that a subset of the attributes and relational structure from the source and target domains are imported into a blended space where they can be combined and supplemented with information from background knowledge (Coulson, 1996, 2000). These hybrid models, or *blends*, are useful in explaining discrepancies between the way that shared representations function in the source and target domains, as well as emergent properties evoked by metaphoric expressions (Tourangeau & Rips, 1991). For example, blending theory explains why it is insulting to call a surgeon a *butcher*, even though meat cutters are not customarily considered incompetent. Grady, Oakley, and Coulson (1999) suggest that the incompetence inference arises from the composition of the butcher's techniques and instruments with the surgeon's goals in the blend. In the blend, the hybrid surgeon–butcher performs surgery on a human in the same manner a butcher might operate on a cow carcass. This unpleasant juxtaposition is the origin for the abstract notion of a butcher as someone who uses coarse methods for a job that requires finesse.

Blending theory suggests that metaphor taxes the comprehension system for two reasons: First, it involves the establishment of mappings between elements in distantly related domains, and second, it often requires the activation of background knowledge for information from the two domains to be integrated. However, neither of these operations is unique to metaphor comprehension. Conceptual blending theory suggests that all language comprehension involves the construction of multiple cognitive models and the establishment of mappings between their components. For example, in the literal use of *gem* in (1), the reader must establish a mapping between *the stone we saw in the natural history museum* and *a gem*, on the basis of category membership.

(1) That stone we saw in the natural history museum is a gem.

Table 1
Examples of the Experimental Sentences

Literal: *He knows that whiskey is a strong intoxicant.*
Literal mapping: *He has used cough syrup as an intoxicant.*
Metaphor: *He knows that power is a strong intoxicant.*

Literal: *The carnival featured an orangutan, a sword swallower, and even a cannibal.*
Literal mapping: *He wondered why the African tribesman was portrayed as a cannibal.*
Metaphor: *She was sexy, but he'd heard she was a real cannibal.*

Literal: *The secret ingredient in her stew is cayenne.*
Literal mapping: *The chef apparently uses salt instead of cayenne.*
Metaphor: *My crazy uncle says jokes are conversation's cayenne.*

Literal: *They had a few chickens in the yard, and in the barn was a goat.*
Literal mapping: *On our last trip into the mountains, Dad thought a bighorn sheep was a goat.*
Metaphor: *Someone had to take the fall, and unfortunately your husband was the goat.*

Literal: *Turns out, it wasn't any rare species of insect, just a flea.*
Literal mapping: *Some subjects got the disease from a mosquito instead of a flea.*
Metaphor: *The independent prosecutor thought he was a bulldog, but he was really more of a flea.*

Literal: *They just announced that the governor was charged with grand larceny.*
Literal mapping: *What I thought was petty theft, the judge thought was grand larceny.*
Metaphor: *I knew she was out to steal his heart, but that kiss was grand larceny.*

Literal: *The U.N. committee found evidence of widespread malnutrition.*
Literal mapping: *He mistook their crowd's stylish look for malnutrition.*
Metaphor: *He complained that prison life was spiritual malnutrition.*

Literal: *He knew he'd have to work all night, so the last thing he needed was a headache.*
Literal mapping: *The doctor diagnosed his tumor as a headache.*
Metaphor: *The actor says interviews are always a headache.*

Literal: *The conductor had no idea the train had been boarded by a known villain.*
Literal mapping: *In the best part of the movie, the hero has to impersonate the villain.*
Metaphor: *Many people in the agency now believe that plastics are an environmental villain.*

Literal: *I read that one of Canada's major exports is maple syrup.*
Literal mapping: *In the movie Psycho, the blood was really cherry syrup.*
Metaphor: *He didn't understand the words, but her voice was sweet syrup.*

Comprehension of the grammatically cued mapping in this literal sentence can be achieved with fairly minimal retrieval and comparison of conceptual structure associated with the two objects in question.

In contrast, the metaphoric use of *gem* in (2) appeals to particular abstract and relational aspects of the reader's concept of gems.

(2) After giving it some thought, I realized the new idea was a gem.

In (2), the speaker's idea is mapped analogically onto the concept of a gem. Only some of a gem's typical qualities are imported into the new blended space in which jewels and thoughts overlap, and these qualities are related to a real gem's properties only analogically. Although *clarity* in a gem refers to the unimpeded passage of light, *clarity* in an idea refers to the unimpeded passage of knowledge.

Similarly, both gems and ideas can be *beautiful*, but standards of attractiveness are qualitatively different. In the blending model, such mappings are made possible because of the incorporation of background knowledge, which sometimes includes underlying conceptual metaphors.

In the corresponding *literal-mapping* condition in (3) the use of *gem* is fully literal but involves fairly extensive mapping between the pebble in the tin ring and the gem in a more prototypical ring.

(3) The ring was made of tin, with a pebble instead of a gem.

Some common properties of pebbles and gems—shape, size, and hardness—allow them to fill the same slots in the relational structure of a ring. Successful mapping involves one's understanding that a pebble can top a toy ring, just as a gem can top a piece of fine jewelry, while discounting noncorresponding properties of pebbles and gems that are irrelevant (expense, rarity, brightness, etc.). We suggest that such cases are intermediate between fully literal and clearly metaphoric uses. Like other literal uses, literal mappings appeal to the literal meaning of the term and invoke concrete attributes of the relevant concepts. But like metaphors, their comprehension requires the apprehension of mappings between two cognitive models.

Our literal mapping sentences include contexts in which one thing is substituted for another, mistaken for another, or used to represent another in child's play, drama, or deception (see additional examples in Table 1). Disparate though these examples may be, they all require the reader to recognize the similarities and differences between two cognitive models as in true metaphors like (2). When one uses a chair instead of a ladder, for example, it is important to understand that one can stand on a chair (as well as sit in it) and that it is possible to reach elevated heights when standing on a chair, just as it is with a ladder. When a boy in a sheet represents a ghost, it is important to understand that he shares some attributes of a ghost (e.g., being white), as well as some relations (he scares other children participating in the game). Our prediction is that comprehension of these literal mappings, like the comprehension of metaphors, will mandate an evaluation of the correspondence between two cognitive models and the selection and alignment of some shared attributes and relations.

Thus, in the present study, we used triplets of sentences hypothesized to fall on a gradient of processing difficulty, from literal statements of class inclusion as in (1), to literal mappings as in (3), to the fully metaphoric uses as in (2). We note, however, that although blending theory provides a ready definition of literal mappings as falling midway between literal and metaphoric language, it is quite possible that other models of metaphor comprehension would provide convergent definitions.

Event-Related Brain Potentials

The second relatively innovative aspect of the present study is that we recorded event-related brain potentials

(ERPs), a record of synaptic potentials that are synchronized to stimulus presentation (see Rugg & Coles, 1995, for a review). Quantitative differences in neurophysiological processes are indexed by ERPs that have the same polarity, waveshape, and scalp distribution, but differ in amplitude or latency. Qualitative differences are indexed by ERPs that differ in polarity, waveshape, and scalp distribution. Although both total reading and lexical decision times have suggested that times to comprehend metaphoric and literal statements do not differ, measurement of ongoing brain activity might either indicate more effortful comprehension or detect a qualitative difference in metaphoric and literal comprehension mechanisms.

A second general motivation for using a neurophysiological measure is that neuropsychology provides the one bit of evidence that has not been well accommodated by the continuity claim assumed in most contemporary models of metaphor comprehension. In contrast to the aphasias associated with left hemisphere damage, more subtle communicative deficits are observed after right hemisphere strokes, one of which has been characterized as difficulty understanding nonliteral language (Brownell, Potter, & Michelow, 1984; Brownell, Simpson, Bihrle, & Potter, 1990; Winner & Gardner, 1977). If indeed right hemisphere damage can selectively impair the comprehension of nonliteral language, this bolsters the standard model's claim that figurative language requires qualitatively different processing mechanisms than does "normal" language. Because laterally asymmetric ERPs are commonly observed in both perceptual and psycholinguistic studies (see King, Ganis, & Kutas, 1998, for a review), ERPs might provide a good measure of the differential contribution of the two cerebral hemispheres to processing metaphoric language.

In the present study, one ERP component of particular interest is the N400 (N for its negative polarity, and 400 for its peak latency at 400 msec after the onset of the stimulus). All words elicit N400, and the amplitude of this component indexes the ease or difficulty of semantic integration in literal sentences (see Brown & Hagoort, 1994; Kutas, Federmeier, Coulson, King, & Muente, 2000; Kutas & Van Petten, 1994, for reviews). For sentence-final words, N400 amplitude is inversely related to cloze probability, an off-line measure of semantic constraint (Kutas & Hillyard, 1984). For sentence-intermediate words, N400 is large at the beginning of a sentence, particularly for low-frequency words, but declines with increasing semantic constraints as a sentence proceeds (Van Petten, 1995). Our a priori prediction was that the N400 component of the ERP would show graded amplitudes across the literal, literal mapping, and metaphor conditions, reflecting a concomitant gradient of processing difficulty.

The present design provides a partial replication and extension of a study by Pynte, Besson, Robichon, and Poli (1996). Those investigators compared ERPs elicited by final words of familiar French metaphors like *Those fighters are lions*, unfamiliar metaphors like *Those apprentices are lions*, and literal statements of class inclusion like

Those animals are lions. Although the same set of words was used and they were matched on cloze probability, familiar metaphoric endings elicited larger N400s than did literal endings, and unfamiliar metaphors elicited larger N400s than did familiar metaphors. However, no behavioral data were collected for the unfamiliar metaphors, and it is possible that some of these were not correctly interpreted by the participants, but read as literal incongruities.

Because the present study was conducted in order to compare the processing difficulty of literal, literal mapping, and metaphoric sentences, a critical aspect of experimental design was to ensure that all three sentence types were equally interpretable. Three steps were taken to ensure that none of the stimuli were perceived as semantically anomalous and that they were indeed interpreted correctly. First, the metaphors were embedded in sentences that supplied some context (as opposed to the simple *some xs are ys* format sometimes used in metaphor research). Second, the three sentence types were subjected to a cloze procedure in which the participants predicted the final words on the basis of the sentence frames. The final stimulus set was selected so that the same final words were offered equally often as completions of literal, literal mapping, and metaphoric sentences by a normative group. Third, each sentence in the experiment was followed by a comprehension question, and only those accompanied by correct answers were included in the data set.

If processing difficulty is related to the difficulty of mapping and integration, we should observe a gradient of N400 amplitude that reflects the hypothesized mapping and integration difficulty in literal, literal mapping, and metaphoric uses of the same set of words. However, the continuity thesis would be falsified if metaphors elicited ERPs with a different scalp distribution, such as being differently lateralized than ERPs in the literal conditions.

METHOD

Participants

Eighteen native English speakers (14 men, 4 women) were paid for their participation. Their average age was 26 years (range 21–34). Five were left-handed, and 5 were right-handers who reported familial sinistrality. All had normal visual acuity; none had any reported history of neurological or psychiatric disorders. The participants were given the reading span test of working memory (Daneman & Carpenter, 1980), and the Peabody Picture Vocabulary Test (PPVT–R; Dunn & Dunn, 1981). Data were collected from 3 additional participants, but not analyzed: 2 participants displayed excessive eye movement artifacts, and 1 had test scores that suggested a learning disability (77 on PPVT-R and 1.5 on the reading span test, as compared with means of 119 [$SD = 9.7$] and 3.5 [$SD = 0.93$] for the rest of the participants).

Materials

The experimental materials included 165 triplets like those in Examples 1–3 above, in which the same word was used literally, metaphorically, or in the literal mapping condition. More triplets are shown in Table 1. Prospective sentence frames were given to at least 80 people from the University of Arizona community in a cloze task. Mean cloze probabilities were equal (3%; range 0%–88%) across the three conditions, as were sentence lengths (12 words; range

Table 2
Sample Sentences and Their Comprehension Questions

Amidst all the trappings of success, his wife was his anchor.
His wife held him back and kept him from enjoying life. (True/False)

Once suffused with hope, the priest had become a broken vessel.
The priest had lost some of his youthful idealism. (True/False)

Tony knew he'd blown it when he mistook his boss's wife for his mistress.
Tony got confused between his own wife and his mistress. (True/False)

He pretended the soup was a narcotic.
*He wanted to believe the soup would calm him down and make him
  feel better.* (True/False)

The secret ingredient in her stew is cayenne.
It's the spices that make her stew special. (True/False)

She's tired of his continual grumbling.
She doesn't mind his constant complaining. (True/False)

5–19). The triplets were divided into three lists, each consisting of 55 literals, 55 literal mappings, and 55 metaphors, so that while a given participant saw each critical (sentence-final) word in only one of its three possible sentence frames, each word occurred in every condition across participants.

Procedure

The sentences were presented one word at a time, for a duration of 200 msec each. The interword interval was length dependent: 100 msec plus an additional 37 msec for each character in the word. Sentence-final words were presented for 200 msec, with a 2,600-msec period before the onset of the true/false question. Table 2 includes examples of the comprehension questions. In contrast with the word-by-word presentation of experimental stimuli, comprehension questions were presented in their entirety for free reading. The questions were displayed for 6 sec, and the participants responded true or false via a buttonpress (response hands were counterbalanced across subjects). Accuracy on these questions was encouraged over speed. After each question, there were 2 sec of blank screen before the beginning of the next trial.

After the presentation of experimental stimuli, the participants were asked to perform a pencil-and-paper task of rating each sentence for its metaphoricity. The scale ranged from 1 to 5, where 1 was *very literal*, 2 was *somewhat literal*, 3 was *not sure*, 4 was *somewhat metaphoric*, and 5 was *very metaphoric*. Mean ratings for literal, literal mapping, and metaphor stimuli were 1.4, 1.9, and 4.4, respectively. The metaphor stimuli were thus rated as more metaphoric than were literals [$F(1,17) = 612.0, p < .0001$] and literal mappings [$F(1,17) = 451.7, p < .0001$]. Literal mappings were rated as more metaphoric than were literals [$F(1,17) = 38.9, p < .001$]. The low metaphoricity rating of the literal mappings indicates that they were largely interpreted as literal statements, although they were less likely to be rated as *very literal* than were the literals (55% vs. 74% of items, respectively).

Electrophysiological Recording

The electroencephalogram (EEG) was recorded with tin electrodes mounted in a commercially available elastic cap. Midline frontal (Fz), central (Cz), and parietal (Pz) recording sites were used, along with lateral pairs of electrodes over parietal (P3 and P4) and occipital (O1 and O2) scalp as defined by the 10–20 system (Jasper, 1958). Three lateral pairs were also used: (1) a frontal pair placed midway between F7–8 and T3–4 (approximately over Broca's area and its right hemisphere homologue, BL and BR, respectively), (2) a temporal pair placed 33% lateral to Cz (TL and TR), (3) a temporoparietal pair placed 30% of the interaural distance lateral and 12.5% of the nasion–inion distance posterior to Cz (approximately over Wernicke's area and its right hemisphere homologue, WL and WR, respectively). Each scalp site was referred to the left mastoid on line

and later re-referenced to an average of the left and right mastoid sites. The electrodes were also placed under the right eye and at the outer canthi to monitor blinks and eye movements.

The EEG was amplified by a Grass Model 12 polygraph with half-amplitude cutoffs of 0.01 and 30 Hz, digitized on line with a sampling rate of 170 Hz and stored on disk for subsequent averaging. Trials with eye movement, muscle, or amplifier blocking artifacts were rejected off line prior to averaging. This resulted in the rejection of an average of 26% of all trials. ERPs were timelocked to the onset of sentence-final words in each of the three conditions.

RESULTS

Comprehension

All participants responded correctly to at least 84% of the comprehension questions, with a mean of 91% correct ($SD = 4\%$). A repeated measures analysis of variance (ANOVA) with the factors metaphoricity (three levels) and participants (18) revealed no difference in performance on questions following literal, literal mapping, and metaphoric stimuli ($F < 1$). Sentences followed by incorrect answers were not included in the analyses below.

Event-Related Potentials

Figure 1 displays the ERPs elicited by sentence-final words in each of the three conditions. As in other paradigms using visual words, the ERPs were characterized by an N100 at frontal and central scalp sites, a P100 and N180 at occipital sites, and a broadly distributed P2 component. These were followed by an N400 visible at all scalp sites, followed by a late positivity largest at parietal sites. Metaphors elicited larger N400s than did literal sentences, with literal mappings falling between metaphors and literals. Metaphors elicited a larger late positivity than did the other two conditions at posterior (parietal and occipital) scalp sites. At frontal scalp sites (Fz, Bl, Br), the literal mapping condition elicited the largest late positivity.

In many previous studies in which sentence stimuli have been used, N400 effects are longer in duration than those observed here, often spanning a latency window of 300–700 msec (e.g., Van Petten, 1993). In the present study, visual inspection of the data (and the statistical analyses below) suggest that at least two distinct components of the ERP were modulated by sentence type and that the N400 and a late positivity (or positivities) overlapped in time. We consider the 300–500-msec latency window to provide a relatively pure measure of N400 amplitude and the later windows to reflect primarily the late positivities. In contrast, the 500- to 700-msec latency range is likely to be the time region with maximal overlap between the earlier N400 and the later positivities, providing no clear measure of either. Indeed, analysis of the 500- to 700-msec time window yielded no significant main effect of sentence condition. Consequently, the data were quantified by measuring mean voltages in two time windows: the first in the peak latency range of the N400 (300 to 500 msec after sentence-final word onset) and a second window of 700–1,100 msec that spans the post-N400 positivity.
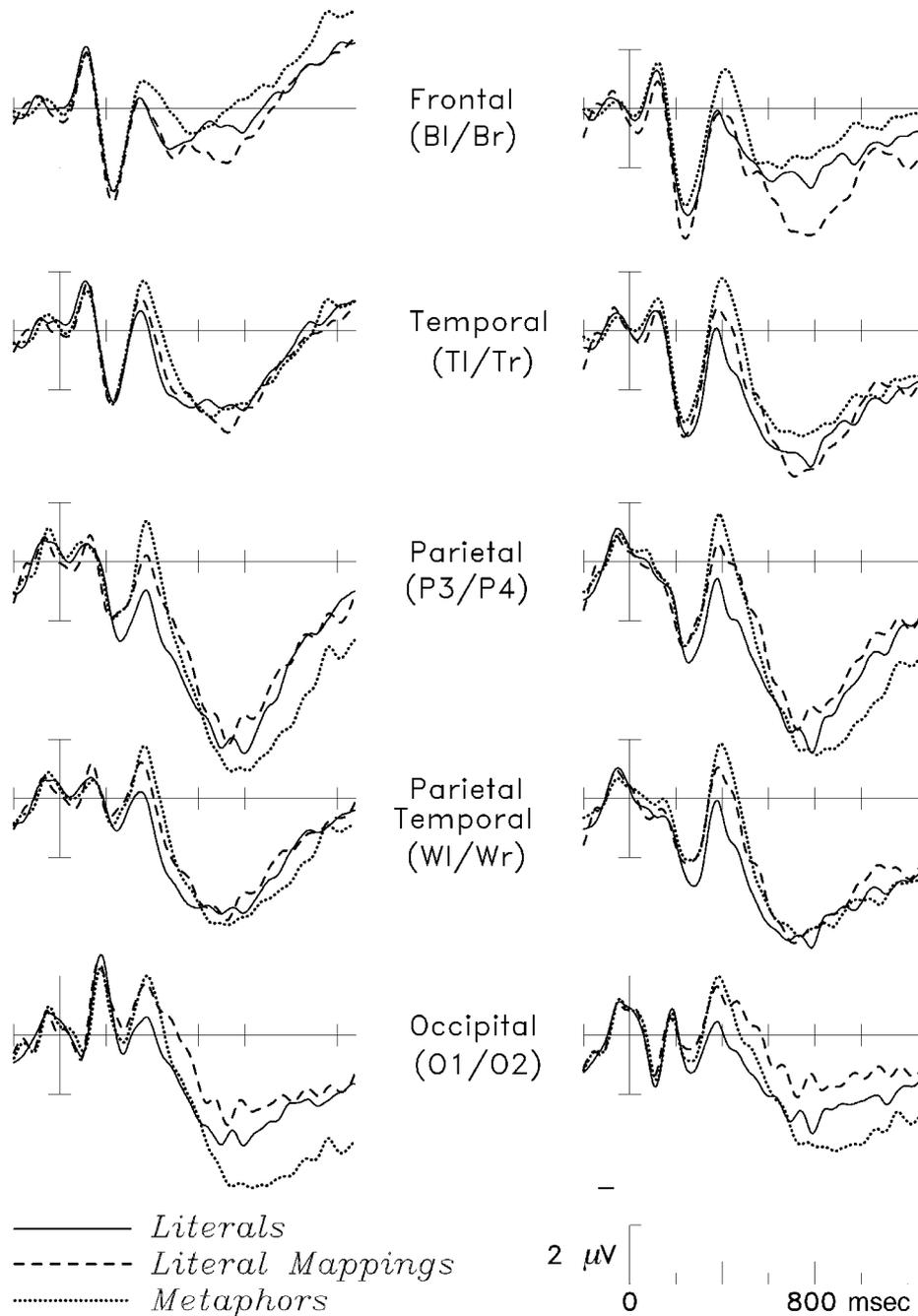
Frontal
(Bl/Br)

Temporal
(Tl/Tr)

Parietal
(P3/P4)

Parietal
Temporal
(Wl/Wr)

Occipital
(O1/O2)

——————— *Literals*
- - - - - - *Literal Mappings*
····················· *Metaphors*

2 μV

0   800 msec

Figure 1. Grand average event-related brain potentials elicited by the sentence-final words, at the lateral scalp sites.

N400. An initial ANOVA of the 300- to 500-msec latency range, with sentence (metaphors vs. literal mappings vs. literals) and scalp site (13 levels) as factors yielded a main effect of sentence [$F(2,34) = 3.90, p <$ .05, $e = .95$].[1] Simple pairwise comparisons showed that the metaphors elicited significantly larger N400s than did the literal statements [$F(1,17) = 6.86, p < .02$]. The lit-

eral mapping condition did not differ significantly from either the metaphors [$F(1,17) = 2.03$] or the literal sentences [$F(1,17) = 2.18$]. This pattern of results is not surprising, since the literal mapping sentences were designed to be a bridge between metaphorical and literal sentences.

An ANOVA with orthogonal trend analysis was used to assess whether N400 amplitudes followed a gradient
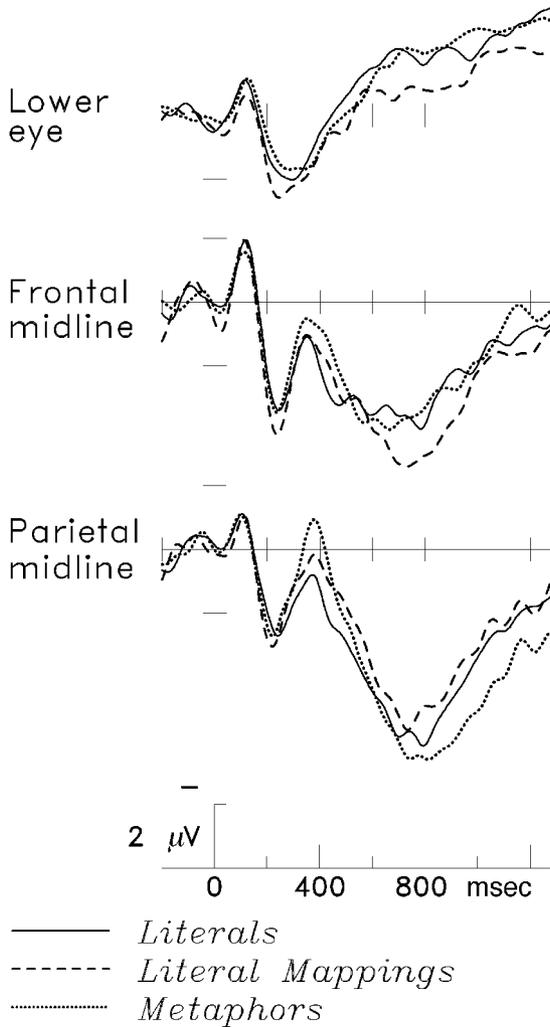
Figure 2. Grand average event-related brain potentials elicited by the sentence-final words, from sites below the left eye, frontal midline (Fz), and parietal midline (Pz).

across the three conditions. In this analysis, the literal, literal mapping, and metaphor conditions were specified as three ordered points (*literal* = 1, *literal mapping* = 2, *metaphor* = 3), rather than as simply three different points as in the standard ANOVA. The gradient of metaphoricity ratings yielded a significant linear effect on N400 amplitude [$F_{linear}(1,17) = 6.86, p < .02$]. The quadratic trend component was not significant ($F < 1$). The trend analysis thus indicates that the gradient of N400 amplitude across the three conditions was robust, although the differences between literal mapping and literal, and between literal mapping and metaphor were small ones. The linear trend across the three sentence types accounted for more than 98% of the total variance due to sentence type. In contrast, a linear trend analysis that stipulated that the literal and literal mapping conditions occupy the same position on a *metaphoricity gradient* (points 1, 1, and 3 for literal, literal mapping, and metaphor, respectively) accounted

for much less of the total variance due to sentence type (i.e., only 40%). These analyses indicate that treating the three conditions as a graded continuum provides a better account of the data[2] than does a theory that stipulates a binary cut dividing the two literal conditions from the metaphor condition.

Although the analyses above revealed no interactions between sentence type and scalp site, the spatial distributions of the condition differences were examined in more detail via analyses of the five pairs of lateral scalp sites, taking metaphoricity (3 levels), anterior to posterior location (AP, 5 levels), and laterality (left vs. right) as factors. This ANOVA yielded a main effect of sentence condition [$F(2,34) = 4.47, p < .05, e = .94$], but no significant interaction between sentence condition and AP [$F(8,136) = 2.09$] and no significant interactions involving hemisphere ($F < 1$).

Late positivities. In contrast to the spatially widespread gradient of N400 amplitude, Figure 1 shows that the late positive difference across conditions had a more restricted scalp distribution: Literal mappings elicited a larger positivity than did the other two conditions at frontal sites, whereas metaphors elicited a larger positivity than did the other two conditions at parietal, parietotemporal, and occipital sites. Differences among the sentence types in frontal and posterior positivities are also shown in Figure 2. An initial analysis of the 700- to 1,100-msec time window with sentence type (3 levels) and scalp site (13 levels) as factors yielded only an interaction of sentence type × site [$F(24,408) = 10.4, p < .001, e = .40$], unaccompanied by a main effect of sentence type. A follow-up analysis of the lateral electrode pairs showed that the interaction between condition and site was driven by sentence differences between the front and back of the head, rather than lateralized differences [sentence × anterior/posterior, $F(8,136) = 16.8, p < .001, e = .45$; sentence × hemisphere, $F(2,34) = 1.48$].

Significant interactions between sentence type and scalp site afford two general sorts of interpretations. On the one hand, such interactions might reflect an amplitude modulation of an ERP component that is present in all experimental conditions and that is always larger at some scalp sites than at others. For instance, doubling the strength of a single hypothetical cortical "source" will similarly produce multiplicative changes in amplitude across scalp locations—numerically large changes at sites with large initial amplitudes and smaller changes at sites with small initial amplitudes. Because the ANOVA uses an additive rather than a multiplicative model, such changes in amplitude might yield condition × site interactions, although the spatial distribution of the component is identical across conditions (McCarthy & Wood, 1985). On the other hand, true changes in scalp distribution across conditions might also be indexed by condition × site interactions, and it is of some interest to detect such changes because they reflect the activity of different populations of neurons across conditions. In order to discriminate between these two interpretations, we used a normalization proce-

dure that eliminates overall amplitude differences between conditions (McCarthy & Wood, 1985). After normalization, significant condition $\times$ site interactions would indicate genuine differences in the scalp distribution of ERPs elicited by different conditions. Pairwise comparison of normalized measures from the literal and metaphor conditions yielded no such significant interaction [sentence $\times$ anterior/posterior, $F(4,68) = 2.03$]. This result suggests that the posterior positivity elicited by metaphors is merely an amplitude enhancement of the posterior positive component present for all three sentence types.

In contrast, comparisons of the literal mapping condition and each of the other two sentence types did yield significant interactions between sentence type and the anterior/posterior factor after normalization [literal mapping vs. literal, $F(4,68) = 6.10, p < .01, e = .46$; literal mapping vs. metaphor, $F(4,68) = 16.9, p < .001, e = .44$]. The latter results indicate that the large frontal positivity was distinctive of the literal mapping condition. Figure 1 suggests that the frontal positivity was slightly larger over the right than the left, reflected in a three-way interaction of the factors of sentence type, anterior/posterior, and hemisphere in the comparison of literal mapping with literal sentences [$F(4,68) = 3.24, p < .05, e = .58$], although not in the comparison of literal mappings with metaphors.

Figure 2 suggests that the posterior positivity was not only largest for metaphors, but also of slightly longer latency. Across the posterior sites (Pz, P3, P4, Wl, Wr, O1, O2), the positive component reached peak amplitude at 804 msec ($SE = 9$) for literal sentences, 819 msec ($SE = 10$) for literal mappings, and 845 msec ($SE = 10$) for metaphors [$F(2,34) = 3.41, p < .05, e = .95$]. Pairwise comparisons showed that the latency shift between literal and metaphor sentences was significant [$F(1,17) = 5.62, p < .05$], whereas the literal mapping condition did not differ significantly from either of the other two conditions (much like the results for the N400 amplitude described above). An ANOVA with orthogonal trend analysis showed that the gradient of peak latency across the three sentence conditions was linear with respect to the mean metaphoricity ratings offered by the participants [$F_{\text{linear}}(1,17) = 4.73, p < .05$].

DISCUSSION

The results confirmed our central prediction of graded N400 amplitude across sentence-final words used literally, metaphorically, and in the intermediate literal mapping condition. Because N400 amplitude has generally been correlated with factors suspected to increase semantic processing difficulty (i.e., weak or absent semantic context, presence of low-frequency words), we interpret this finding as indicating a gradient of difficulty in sentence comprehension across the three conditions. The N400 amplitude difference between literal and metaphoric sentences replicates that reported by Pynte et al. (1996). As in that study, the N400 difference between literals and

metaphors observed here was rather small. The absolute magnitude of the literal/metaphor difference is most comparable with that previously observed in comparisons of sentence-final words with a cloze probability discrepancy of some 20%, or between high- and low-frequency words in the absence of semantic context (Kutas & Hillyard, 1984; Van Petten, 1993). But given that the literal/metaphor difference cannot be attributed to either cloze probability or word frequency, we conclude that it was more difficult for readers to process the metaphors. Also, as in Pynte et al.'s (1996) study, the N400 elicited by metaphoric and literal words was not differentially lateralized, despite reports that right hemisphere damage specifically impairs the comprehension of nonliteral language (Brownell et al., 1990).[3]

The novel finding here was the identification of a sentence type that behaved midway between frankly metaphorical and transparently literal. The gradient of N400 amplitude is consistent with the continuity claim that literal and metaphoric language share some processing mechanisms, but inconsistent with the equivalence claim that comprehension of metaphoric language is no more effortful than literal language.[4] These findings, then, raise the question of the nature of the processing difficulty. Namely, what made both the metaphoric and literal mapping sentences more "difficult" so that they yielded enhanced N400s? One prominent psycholinguistic model of metaphor comprehension—Glucksberg's *property attribution model*—has little to say on this point. In that model, metaphors are read as statements of class inclusion, so that the *shark* in *My lawyer is a shark*, refers to a class of predatory creatures that also includes the speaker's lawyer. The source domain in this model is an abstract superordinate category that has not yet been lexicalized (e.g., *things that are vicious and aggressive*), and successful metaphor comprehension consists of attributing the properties of this category to the target term (Glucksberg, 1998; Glucksberg & Keysar, 1990). The sentences in all three of the conditions used here can be read as class inclusion statements of the sort described by Glucksberg. Even for the literal sentences, the "source" terms were rarely lexicalized categories (e.g., *furniture* or *animals*), but were more often complex propositions such as *a major export of Canada* or *the last thing one needs when working all night* (see Table 1). By itself, a definition of metaphor as a class inclusion statement does not explain the gradation of difficulty indexed by the graded amplitude of the N400 across literal, literal mapping, and metaphoric sentences.

As noted above, blending theory suggests that metaphor taxes the comprehension system because it involves (1) the establishment of mappings between elements in distantly related domains, and (2) the retrieval of information from memory to integrate these elements. Consequently, we attribute the enhanced N400 in both the literal mapping and metaphor conditions to the fact that they both include an invitation to discover the similarity between two entities and that the similarity is only partial. We suggest that ini-

tial semantic conflicts between source and target domains are responsible for the larger N400s in both the metaphor and literal mapping conditions. The differential N400 across conditions might thus arise during an early stage of comparison between source and target terms that might correspond to *alignment* in Gentner and colleagues' model or to *mapping* in conceptual blending theory (Coulson, 2000; Gentner & Wolff, 1997; Wolff & Gentner, 2000).

In addition to larger N400s, metaphors also elicited a larger and later positivity at posterior scalp sites than did either literal or literal mapping sentences, which did not differ from each other.[5] Although the latency of this positive peak was a graded function of figurativity (shortest for literals, longest for metaphors, with literal mappings falling in between), the amplitude of this positive peak was specifically sensitive to metaphors. This finding is consistent with the continuity claim, since the literal condition elicited a positivity with the same distribution across the scalp, only smaller in amplitude and earlier in peak latency.

Moreover, literal sentence-final words have occasionally been observed to elicit a positive peak after the N400 in previous ERP studies. Little has been written about the psychological factors affecting this sentence-ending positivity, although its intermittent presence suggests that it is dissociable from the N400 and reflects different cognitive operations.

In published studies, the only factor that has reliably influenced the amplitude of the sentence-ending positivity is word frequency. With weak semantic support, low-frequency words simply elicit larger N400s than do high-frequency words (Van Petten, 1993). But when they serve as semantically predictable sentence completions, low-frequency words elicit a larger posterior positivity than do high-frequency words. Van Petten, Kutas, Kluender, Mitchiner, and McIsaac (1991) have suggested that the word-frequency effect for the sentence-ending positivity reflects a difference in the lexical semantics of high- and low-frequency words, specifically that the more detailed and precise meanings of low-frequency words (Zipf, 1945) mandate more extensive retrieval of information from semantic memory in the course of arriving at a sentence-level interpretation.

The sensitivity of the posterior positivity to word frequency suggests a possible interpretation for the positivity observed here for metaphoric sentence completions. Although the metaphoric and literal endings were identical in orthographic form, their comprehension required retrieval of different aspects of conceptual structure. Given that the metaphors were relatively novel, the relevant concepts were unlikely to have been strongly associated with the orthographic form of the word, but instead required the recovery and integration of additional material from semantic memory,[6] including conceptual metaphors of the sort described by Lakoff (1993). A search for such information might be triggered by the initial semantic mismatches indexed by the N400; successful retrieval of the relevant conceptual metaphor (indexed by the pos-

terior positivity) would then provide the necessary bridge between the distantly related source and target terms and allow the appropriate blended concepts to be constructed.

REFERENCES

ARISTOTLE. (1952). *Aristotle* (Great Books of the Western World, Vol. 8). Chicago: Encyclopedia Britannica.

BLASKO, D. G. (1999). Only the tip of the iceberg: Who understands what about metaphor? *Journal of Pragmatics,* **31**, 1675-1683.

BLASKO, D. G., & CONNINE, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **19**, 295-308.

BROWN, C., & HAGOORT, P. (1994). Brain response to lexical ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45-80). Hillsdale, NJ: Erlbaum.

BROWNELL, H., POTTER, H., & MICHELOW, D. (1984). Sensitivity to lexical denotation and connotation in brain-damaged patients: A double dissociation? *Brain & Language,* **22**, 253-265.

BROWNELL, H., SIMPSON, T., BIHRLE, A., & POTTER, H. (1990). Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia,* **28**, 375-383.

COULSON, S. (1996). The Menendez brothers virus: Analogical mapping in blended spaces. In A. Goldberg (Ed.), *Conceptual structure, discourse, and language* (pp. 143-158). Palo Alto, CA: CSLI.

COULSON, S. (2000). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction.* Cambridge and New York: Cambridge University Press.

COULSON, S., VAN PETTEN, C., & FOLSTEIN, J. (2000). Event-related brain response to literal and figurative language in left handers. *Psychophysiology,* **37** (Suppl. 1), 532.

DANEMAN, M., & CARPENTER, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior,* **19**, 450-466.

DUNN, L., & DUNN, L. (1981). *Peabody picture vocabulary test—revised.* Circle Pines, MN: American Guidance Service.

FAUCONNIER, G., & TURNER, M. (1998). Conceptual integration networks. *Cognitive Science,* **22**, 133-187.

FLETCHER, P. C., HAPPE, F., FRITH, U., & BAKER, S. C. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition,* **57**, 109-128.

FOSS, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology,* **14**, 590-607.

FRISSON, S., & PICKERING, M. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor & Symbol,* **13**, 149-171.

GENTNER, D., & WOLFF, P. (1997). Alignment in the processing of metaphor. *Journal of Memory & Language,* **37**, 331-355.

GERNSBACHER, M. A., & ROBERTSON, R. (1999). The role of suppression in figurative language comprehension. *Journal of Pragmatics,* **31**, 1619-1630.

GIBBS, R. W. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **16**, 56-66.

GIBBS, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding.* Cambridge: Cambridge University Press.

GIBBS, R.W., BOGDANOVICH, J. M., SYKES, J. R., & BARR, D. J. (1997). Metaphor in idiom comprehension. *Journal of Memory & Language,* **37**, 141-154.

GILDEA, P., & GLUCKSBERG, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning & Verbal Behavior,* **22**, 577-590.

GLUCKSBERG, S. (1998). Understanding metaphors. *Current Directions in Psychological Science,* **7**, 39-43.

GLUCKSBERG, S., GILDEA, P., & BOOKIN, H. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning & Verbal Behavior,* **21**, 85-98.

GLUCKSBERG, S., & KEYSAR, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review,* **97**, 3-18.

GLUCKSBERG, S., McGLONE, M. S., & MANFREDI, D. (1997). Property

attribution in metaphor comprehension. *Journal of Memory & Language, 36*, 50-67.

GRADY, J., OAKLEY, T., & COULSON, S. (1999). Conceptual blending and metaphor. In R. Gibbs (Ed.), *Metaphor in cognitive linguistics* (pp. 101-124). Philadelphia: John Benjamins.

GRICE, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41-58). New York: Academic Press.

INHOFF, A. W., LIMA, S. D., & CARROLL, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition, 12*, 558-567.

JASPER, H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography & Clinical Neurophysiology, 10*, 371-375.

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

KING, J., GANIS, G., & KUTAS, M. (1998). Potential asymmetries in language comprehension: In search of the electrical right. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience* (pp. 187-213). Mahwah, NJ: Erlbaum.

KISS, G. R., ARMSTRONG, C., MILROY, R., & PIPER, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Atkin, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153-165). Edinburgh: University Press.

KUTAS, M., FEDERMEIER, K. D., COULSON, S., KING, J. W., & MUENTE, T. F. (2000). Language. In J. T. Cacioppo, L. G.Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 576-601). New York: Cambridge University Press.

KUTAS, M., & HILLYARD, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163.

KUTAS, M., & VAN PETTEN, C. (1994). Psycholinguistics electrified. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83-103). San Diego: Academic Press..

KUTAS, M., VAN PETTEN, C., & BESSON, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography & Clinical Neurophysiology, 69*, 218-233.

LAKOFF, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 202-251). Cambridge: Cambridge University Press.

LAKOFF, G., & JOHNSON, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

McCARTHY, G., & WOOD, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography & Clinical Neurophysiology, 62*, 203-208.

ORTONY, A., SCHALLERT, D., REYNOLDS, R., & ANTOS, S. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning & Verbal Behavior, 17*, 465-477.

PYNTE, J., BESSON, M., ROBICHON, F., & POLI, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain & Language, 55*, 293-316.

QUINTILLIAN. (1933). *The Institutio Oratoria of Quintillian*. (H. E. Butler, Trans.). New York: G. P. Putnam's Sons. (Original work published 1921)

RUGG, M., & COLES, M. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford: Oxford University Press.

SEARLE, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

SHEN, Y. (1999). Principles of metaphor interpretation and the notion of 'domain': A proposal for a hybrid model. *Journal of Pragmatics, 31*, 1631-1653.

SIMPSON, G. B., PETERSON, R. R., CASTEEL, M. A., & BURGESS, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 88-97.

SWEETSER, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

TABOSSI, P. (1991). Understanding words in context. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 1-22). Amsterdam: North-Holland.

TOURANGEAU, R., & RIPS, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory & Language, 30*, 452-472.

TURNER, M. (1991). *Reading minds: The study of English in the age of cognitive science*. Princeton, NJ: Princeton University Press.

VAN PETTEN, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language & Cognitive Processes, 8*, 485-531.

VAN PETTEN, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology, 32*, 511-525.

VAN PETTEN, C., KUTAS, M., KLUENDER, R., MITCHINER, M., & MCISAAC, H. (1991). Fractioning the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3*, 131-150.

VAN PETTEN, C., WECKERLY, J., MCISAAC, H., & KUTAS, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science, 8*, 238-242.

WINNER, E., & GARDNER, H. (1977). The comprehension of metaphor in brain-damaged patients. *Brain, 100*, 719-727.

WOLFF, P., & GENTNER, D. (2000). Evidence for role-neutral initial processing of metaphors. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 529-541.

ZIPF, G.K. (1945). The meaning–frequency relationship of words. *Journal of General Psychology, 33*, 251-256.

NOTES

1. Huyhn-Feldt correction for nonsphericity of variance. For all F values with more than one degree of freedom in the numerator, we report the original degrees of freedom, the corrected probability level, and the epsilon correction factor.

2. The mean metaphoricity ratings offered by the participants (1.4 for literals, 1.9 for literal mappings, and 4.4 for metaphors) were also used to specify the ordering of the three conditions in a trend analysis. This analysis also yielded a significant linear trend [$F_{linear}(1,17) = 7.31, p < .02$], although it did not capture as much of the total variance due to sentence type (79%) as did the simple 1–2–3 spacing reported in the text.

3. Note, however, that we did not observe the right-greater-than-left asymmetry typically associated with the N400. We attribute the symmetric topography of these ERPs to the fact that our participants included 5 people with familial sinistrality, a group known for its laterally symmetric N400s (Kutas, Van Petten, & Besson, 1988), as well as 5 left-handers. Although the impact of handedness on ERPs to figurative language is an interesting topic in its own right (Coulson, Van Petten, & Folstein, 2000), handedness of the participants is orthogonal to the within-subjects comparisons that are the focus of the present study.

4. We examined a potential confound for interpreting the gradient of N400 amplitudes across conditions. Although cloze probability was matched across the three conditions, the use of different sentence contexts for identical targets raises the possibility that some sentences included more intermediate words that were lexically associated with the critical final words than others. We searched a large database of free associations (Edinburgh Associative Thesaurus; Kiss, Armstrong, Milroy, & Piper, 1973), which included 93 of the 165 critical target words as responses to cue words. The number of cue (associated) words appearing in the experimental sentences was 16 for the literal, 21 for the literal mapping, and 16 for the metaphor condition, with associative strengths of .11, .03, and .09, respectively. These observations are not consistent with the gradient of N400 amplitudes. Because associative priming effects in sentences are short lived and rapidly attentuated by intervening words (Foss, 1982; Simpson, Peterson, Casteel, & Burgess, 1989; Van Petten, Weckerly, McIsaac, & Kutas, 1997), we also examined a three-word window immediately preceding the critical sentence-final words. In this window, only 15 of the 279 experimental sentences examined included associates of the final words: seven associates with mean strength of .08 for literal, five associates with strength of .08 for literal mapping, and three associates with strength of .11 for metaphors. The small numbers and weak associative strengths between intermediate and final words suggest that this factor had little impact on the observed N400 gradient.

5. In a similar comparison between cloze-matched literal and metaphorical sentences, Pynte et al. (1996, Experiment 1) did not observe a reliably larger late positivity for metaphors. In another experiment in which unfamiliar metaphors with supporting context were used, Pynte et al. did observe a larger late positivity, but the control condition consisted of familiar metaphors with irrelevant context, so that it is difficult to directly compare these results with the present ones.

6. Although the metaphoric and literal sentences proved to differ only quantitatively, the results included one striking qualitative difference among sentence types. The literal mapping sentences were designed to be an intermediate condition (and behaved accordingly in N400 amplitude), but elicited a large frontal positivity distinct from both the literal and metaphor conditions. The frontal positive peak elicited by literal mappings does not resemble any phenomenon in the sentence processing literature to our knowledge, so this finding requires replication and extension. It is worth noting, however, that a substantial proportion (74%) of the literal mapping sentences describe situations of pretense, lying, and mistaken identification, so their comprehension depends on understanding the mental states of actors. One speculation is that the unusual frontal positivity elicited by literal mappings is related to the observation that narratives placing heavy demands on *theory of mind* elicit greater blood flow in prefrontal cortex than do narratives that do not (Fletcher, Happe, Frith, & Baker, 1995).