

EMBODIED SENTENCE COMPREHENSION

Rolf A. Zwaan

Carol J. Madden

Florida State University

To appear in: D. Pecher & R.A. Zwaan (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge University Press.

Address all correspondence to:  
Rolf A. Zwaan  
Department of Psychology  
Florida State University  
Tallahassee, FL 32306-1270  
phone: 850-644-2768  
FAX: 850-644-7739  
zwaan@psy.fsu.edu

## 1.0. Introduction

There are two views of cognition in general and of language comprehension in particular. According to the traditional view (Chomsky, 1957; Fodor, 1983; Pylyshyn, 1986), the human mind is like a bricklayer, or maybe a contractor, who puts together bricks to build structures. The malleable clay of perception is converted to the neat mental bricks we call words and propositions, units of meaning, which can be used in a variety of structures. But whereas bricklayers and contractors presumably know how bricks are made, cognitive scientists and neuroscientists have no idea how the brain converts perceptual input to abstract lexical and propositional representations—it is simply taken as a given that this occurs (Barsalou, 1999).

According to an alternative and emerging view, there are no clear demarcations between perception, action, and cognition. Interactions with the world leave traces of experience in the brain. These traces are (partially) retrieved and used in the mental simulations that make up cognition. Crucially, these traces bear a resemblance to the perceptual/action processes that generated them (Barsalou, 1999) and are highly malleable. Words and grammar are viewed as a set of cues that activate and combine experiential traces in the mental simulation of the described events (Zwaan, in press). The main purpose of this chapter is to provide a discussion of this view of language comprehension. To set the stage for this discussion we first analyze a series of linguistic examples that present increasingly larger problems for the traditional view. Consider the following sentences.

(1) The exterminator checked the room for bugs.

(2) The CIA agents checked the room for bugs.

It is clear that the bug in (1) is not the same as the bug in (2). In other words, “bug” is a homonym. The traditional view has no problem accounting for these two different interpretations of “bug” because it simply assumes that the sentential context disambiguates the homonym such that the correct meaning is selected (although the incorrect meaning may be briefly activated, e.g., Swinney, 1979). Thus, in these cases, one might assume a stronger lexical association between “exterminator” and the “insect” meaning of “bug” than between “CIA agent” and that meaning and vice versa for the “microphone” meaning of bug. The following two sentences already present more of a challenge to the traditional view.

(3) Fred stole all the books in the library.

(4) Fred read all the books in the library.

It is clear that “all the books” means two slightly different things in these two sentences. For example, (3) implies that Fred stole all 12 copies of *War and Peace*, whereas (4) in the most likely interpretation means that Fred read only one copy of *War and Peace*. But both words refer to the same thing: a physical object consisting of written pages, bound together and in a cover. This presents a problem for the traditional view of compositionality according to which concepts have atomic meanings that should remain unchangeable across contexts. However, the traditional view can be amended to account for interpretations such as these. Pustejovsky (1995) has proposed that words have different qualia, that is different interpretations, and that these interpretations are selected by other words in the sentence. For example, a book can be both a physical object and a

source of information. Stealing typically involves physical objects (although one can steal glances, kisses, or ideas) and thus “steal” selects the physical-object quale of “book.” Reading, on the other hand, involves information (even when reading facial expressions, tracks in the snow, or passes in a soccer game), and therefore “read” selects the information-source meaning of “book.” In this sense, the bricklayer metaphor can be extended to that of a child playing with legos (a metaphor used in many linguistics courses). Some pieces, like wheels, have different shapes and different sites for attachment. For example, an axle fits into a flange on the inside of the wheel and a little square block fits on the hub. Similarly, some verbs select one quale of a noun, whereas other verbs will select another.

However, it is not clear whether this lego-extended view of comprehension can account for the following sentences.

(5) John pounded the nail into the wall.

(6) John pounded the nail into the floor.

Here, both sentences use the same verb, and in both sentences, this verb selects the same quale of the noun, “nail”; it is a slender usually pointed and headed fastener designed to be pounded in. What is different in the two sentences is the nail’s orientation. There is nothing in the traditional view to suggest that the nail’s orientation should be part of the comprehender’s mental representation. For example, a common way to represent sentences (5) and (6) is (e.g., Kintsch & van Dijk, 1978):

(7) [POUNDED[JOHN, NAIL]], [IN[NAIL, FLOOR]]

(8) [POUNDED[JOHN, NAIL]], [IN[NAIL, WALL]]

Nothing in these propositional representations says anything about the nail's orientation, yet empirical evidence shows that comprehenders routinely represent this orientation (Stanfield & Zwaan, 2001). A similar point can be made about the shape of objects.

(9) He saw the eagle in the sky.

(10) He saw the eagle in the nest.

According to the traditional view the words in these sentences form the building blocks of meaning out of which the comprehender builds a structure that reflects his or her interpretation of the sentence. But how would this work for sentences (9) and (10)?

Surely the eagle in (9) cannot be the same lego brick as the eagle in (10). In (9) the eagle has its wings stretched out, whereas the eagle in (10) most likely has its wings drawn in. Again, there is empirical evidence that comprehenders are sensitive to these differences (Zwaan, Stanfield, & Yaxley, 2002) and again, this is not predicted by the traditional view. The traditional view also does not seem to have a straightforward account for how the following sentences are interpreted.

(11) Jack looked across the room to see where the whisper/explosion came from.

(12) Jack looked across the valley to see where the whisper/explosion came from.

Clearly, whisper is a better fit for (11) and explosion for (12). But how does the traditional theory account for this? The Merriam-Webster dictionary defines “whisper” as follows: “to speak softly with little or no vibration of the vocal cords especially to avoid being overheard.” Nothing in this definition points directly to the distance over which a whisper can be heard by the human ear. Similarly, nothing in the definition of valley—“an elongate depression of the earth's surface usually between ranges of hills or mountains” according to Merriam-Webster—provides explicit information about the

typical width of a valley. It might be argued that both terms contain information that can be used to infer the respective distances (e.g., “softly”, “overhear,” and “mountain”). But if one looks up “mountain”, for example, the most relevant meaning to be found is "a landmass that projects conspicuously above its surroundings and is higher than a hill." Of course, this does little to alleviate the interpretation problem. This is an example of what has become known as the Chinese Room Problem (Searle, 1980). Given that dictionary meanings are not grounded in perception and action, the comprehender who has to rely on a dictionary is given a perpetual runaround (see also Glenberg, 1997).

Finally, consider the following pair of sentences.

(13) The pitcher hurled the baseball to you.

(14) You hurled the baseball at the batter.

There is nothing different about the intrinsic properties of the baseball in (13) and (14), such as shape or orientation. The only difference between the baseball in (13) and that in (14) is the direction of motion. Although the traditional view would not predict any differences in the mental representations of baseballs formed during comprehension of these two sentences (and may even have trouble explaining it post-hoc in an elegant manner), there is evidence that the direction of motion is incorporated into the representations of the two baseballs, yielding distinct simulations (Zwaan, Madden, Yaxley, & Aveyard, submitted). We will discuss this evidence and other relevant evidence in more detail later.

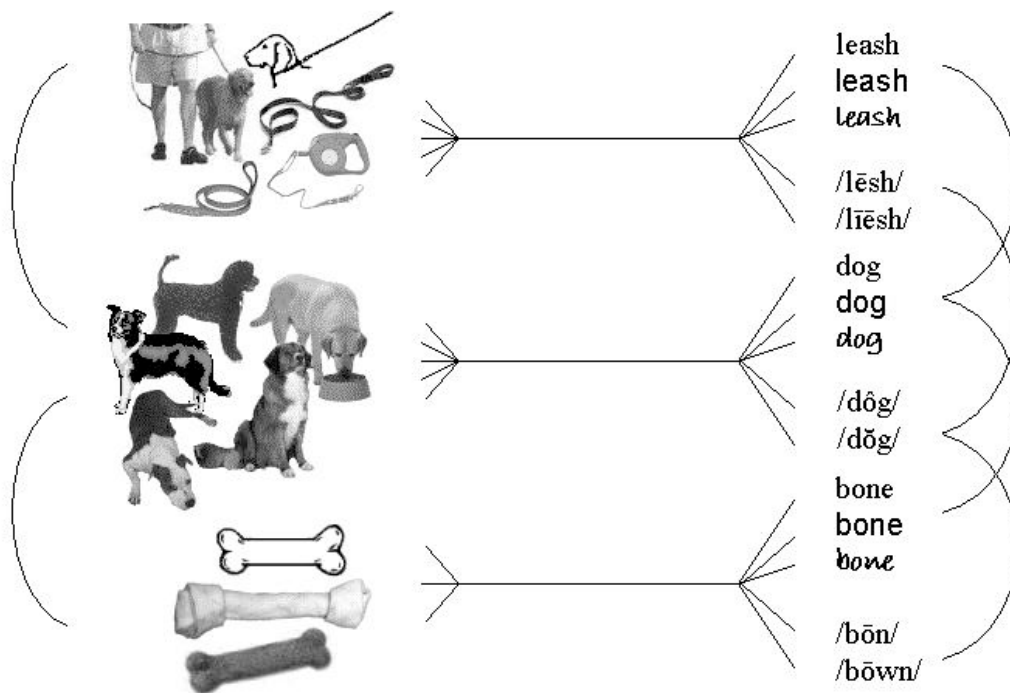
In the rest of this chapter, we propose the beginnings of a theory of sentence comprehension that accounts for these findings in a straightforward way.

## 2. Interconnected experiential traces

In our proposed theory, we assume that all mental representations are experiential. Within the category of experiential representations, we distinguish between referent representations and linguistic representations (see also Sadoski & Paivio, 2001). Referent representations are traces laid down in memory during perceptions of and interactions with the environment. These traces are multi-modal (i.e., combining multiple senses).<sup>1</sup> Because of attentional limitations, these traces are schematic (Barsalou, 1999). The second subcategory of experiential traces consists of linguistic traces. These traces are laid down as linguistic information is being received or produced. For example, there are perceptual traces of hearing, reading, seeing, and feeling (as in Braille) linguistic constructions. As well, there are motor representations of saying, signing, typing, and handwriting linguistic constructions. Not only are these constructions interconnected, they are also connected to referent representations, which are also interconnected (see Figure 1).

---

<sup>1</sup> Damasio (1999, p.160) describes object representations as stored in “dispositional form.” Dispositions are records, which are dormant, rather than active and explicit, as images are. Dispositions include: records of sensory aspects, records of the motor adjustments necessary to gather sensory signals, obligate emotional reaction.



*Figure 1. Schematic depiction of experiential (visual) traces of leashes, dogs, and bones, and of the visual and sound patterns associated with them, as well as the links between these traces.*

How are these interconnections established? The main mechanism is co-occurrence (e.g., Hebb, 1949). Certain entities in the environment tend to co-occur. Ducks are generally found in or near ponds or lakes, monitors on desks, pacifiers in babies' mouths, clouds in the sky, and branches above roots. Events may co-occur or follow in close sequence. For example, a scratchy sound accompanies the action of striking a match, and a flame typically follows it, along with a sulfuric smell, which we may find pleasurable. Because of these spatio-temporal co-occurrences, combinations of entities, actions, events, and bodily states become part of the same experiential trace.

Similarly, linguistic constructs co-occur and may therefore develop associations between themselves. First proposed by associationists such as Hume and Locke, but eschewed by transformational grammarians, the analysis of linguistic co-occurrences has made a recent comeback in the form of sophisticated computational linguistic analyses involving large corpora, such as latent semantic analysis (LSA, Landauer & Dumais, 1997).<sup>2</sup> This idea that associations between representations are formed through co-occurrence of linguistic constructions is central to the current theory. For example, the words “nurse” and “doctor” often co-occur in language. As a result, the experiential traces for these linguistic units are associated, just as the words “peace” and “treaty” and the names Lennon and McCartney. Often sets of more than two words co-occur, as in “hail Mary pass”, “internal revenue service,” and “everything but the kitchen sink.” As a result, entire sequences of words can be treated as constructions (Goldberg, 1995, 2003).

The connections that are established between experiential traces for referents and experiential traces for linguistic constructions are of critical importance to the grounding of language in perception and action (see also Goldstone, Yeng, & Rogosky, this volume). The initial mechanism by which these connections are forged is co-occurrence. When children learn to speak, parents and others point out objects to them in the environment. Moreover, even when children are not attending to the entity in question, they can use mentalistic cues such as the speaker’s eye gazes and facial expressions to form associations between constructs and referents (Bloom, 2000). As a result, children learn to associate an experience of a referent with a particular sound pattern. In fact,

---

<sup>2</sup> These analyses actually go significantly beyond co-occurrences by assessing the degree to which words occur in similar contexts, but this is not relevant to the current argument.

children are surprisingly adept at learning word meaning this way, often needing only a few exposures (Carey & Bartlett, 1978).

Children do not only learn to associate constructs with objects, but also with actions and properties. For example when parents say, “give me the ball,” the child will associate an action— grasping a ball, extending the arm, and then releasing the ball into the grasp of a parent—with a linguistic construction (and with encouraging sounds and facial expressions on the part of the parent). In fact, the child learns something more fundamental, namely that this syntactic construction can be applied in many other contexts—for instance, “throw me the ball,” and even “tell me a story.” As such, the syntactic structure can be thought of as a linguistic construction that conveys meaning (Goldberg, 1995, 2003). The meaning of this construction, the double-object construction, is that an object or something more abstract moves from the agent to a recipient. This is what the different contexts in which the expression is used have in common. Importantly however, this is only part of the meaning of an expression. For example, “throw me the ball” is associated with a different motor program than “give me the ball” and it is also associated with the salient pattern of an object getting smaller in one’s visual field as it moves away from the thrower. On the other hand “tell me a story” is associated with cognitive effort and speech motor programs, as well as with certain encouraging or puzzled facial expressions on the part of the listener. As Hockett (1959) noted, one important feature of language—called displacement—is that it allows us to convey situations that are not part of our immediate environment. The connections between linguistic and referent traces enable this feature. For example, if we have never seen a zebra before and it is described to us as a “horse with black-and-white stripes,”

then we can form a new referent representation by combining the perceptual traces for horses, for stripes, and for black-and-white, based on their associations with the corresponding words (Harnad, 1990). This virtual experiential trace, constructed from a combination of other visual traces can now be stored in long-term memory. Along with it, an association is formed between the sound pattern of “zebra” and the new visual trace. This uniquely human way of learning about the environment through linguistic scaffolding significantly augments what we can learn by interacting directly with the environment.

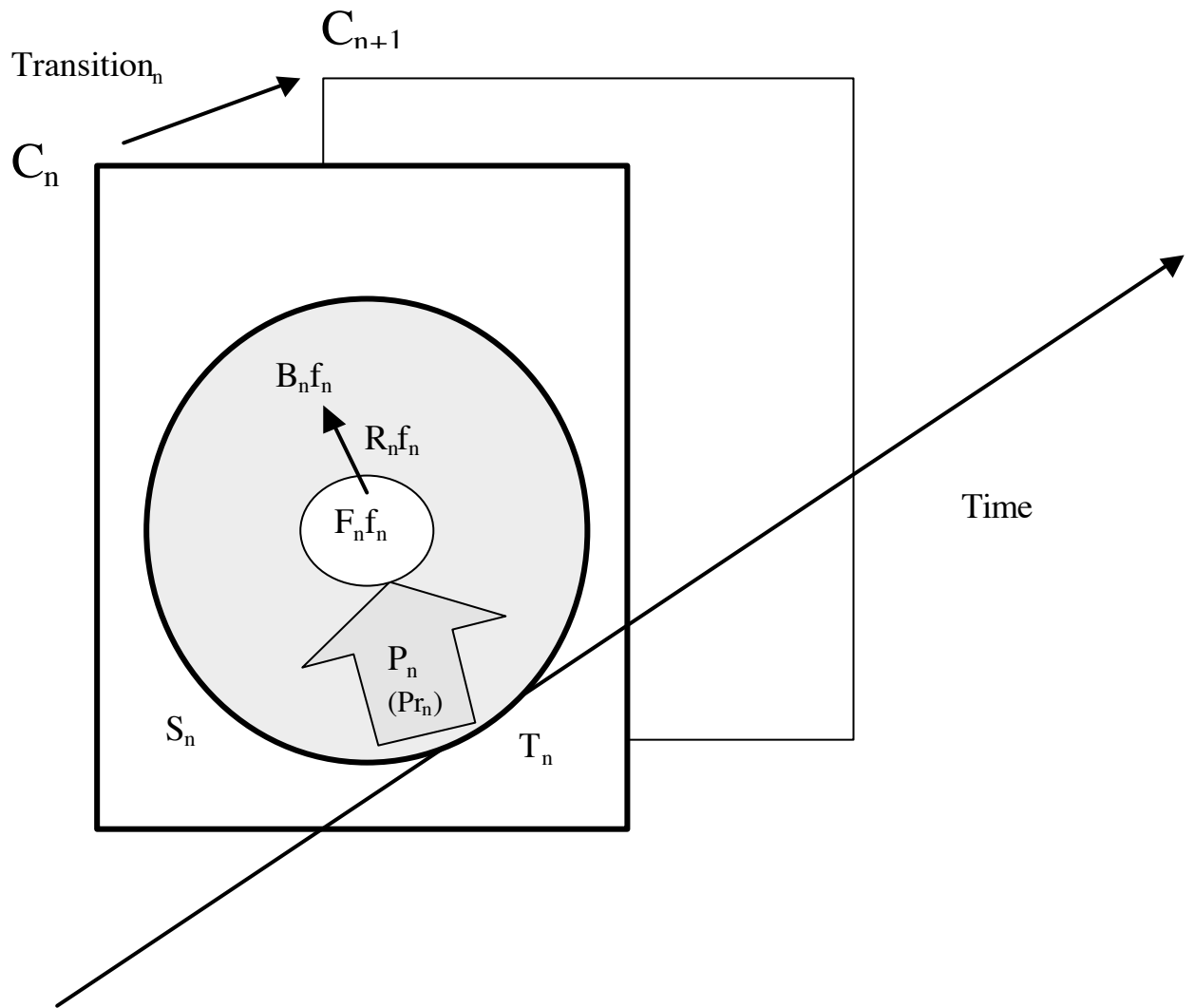
One consequence of the development of connections between the two classes of experiential symbols is that co-occurrences in one domain will produce co-occurrences in the other. These effects are bi-directional. Just as the spatio-temporal proximity of certain objects or events enhances the likelihood that the linguistic constructions denoting them will co-occur, the co-occurrence of linguistic constructions will strengthen the connections between their referents. As a result, although there generally is not an analog first-order mapping between linguistic constructions and their referents, there is a strong second-order mapping. If the link between two constructs is strong, then it is very likely that the link between the corresponding referents is also strong (see Figure 1). This is one reason why techniques such as LSA are often surprisingly successful in capturing meaning. However, as we will show later, many experiential factors are not captured by linguistic co-occurrences.<sup>3</sup>

### 3. Construal

---

<sup>3</sup> There is much more to be said about this issue, but this is beyond the scope of this chapter.

Along with several other researchers, we conceive of language as a set of cues by which the speaker or writer manipulates the listener's or reader's attention on an actual or fictional situation (e.g., Langacker, 1987; Tomasello, 2003). The units in which this process takes place are attentional frames (Langacker, 2001). Attentional frames map onto intonation units, which are speech segments bounded by pauses or intonation shifts (Chafe, 1994). Because written language follows spoken language phylogenetically as well as ontogenetically, the segmentation of spoken language provides the grounding for the segmentation of written language. We define construal as the mental simulation of an experience conveyed by an attentional frame. This mental simulation uses the experiential traces that are activated by the linguistic constructions in the intonation unit.



C = construal

T = time

S = spatial region (personal, action, vista)

P = perspective

F = focal entity

R = relation

B = background entity

f = feature

*Figure 2. The components of construal (from Zwaan, in press).*

Figure 2 (from Zwaan, in press) shows the components of construal. Each construal necessarily includes:

- a time at which the simulated situation occurs (as related to the moment of utterance, to the previously simulated event, and sometimes to some reference point);
- a spatial region in which the described event takes place;
- a perspective (spatial and psychological) from which the situation is experienced;
- a focal entity;
- a background entity.

In addition, the focal entity, relation, and background entity can have features (e.g., size, color, intensity, speed).

Here, our focus is on the process of construal. An important component of construal is the establishment of a focal entity and background entity. Language provides many cues for this, including syntactic, morphemic, and paralinguistic information. In English, for example, the focal entity is usually the first entity described (Langacker, 1987). This means that in active sentences the entity performing an action is the focal entity, whereas in passive constructions the entity undergoing the action is the focal entity. This means that passivization is not without semantic consequences, so that a passive sentence is not really a paraphrase of its active counterpart.

Intonation units typically describe what linguists have traditionally analyzed as events, processes, and states (e.g., Vendler, 1967; Ter Meulen, 1995). Viewing language comprehension as the modulation of attention compels us to treat each as a type of event. Consider the following sentences.

- (15) The car pulled out of the driveway.
- (16) The car was zooming along on the interstate.
- (17) The car was blue.

Punctate events such as (15) are perhaps the most easily conceptualized as mental simulations. The focal entity (the car) initiates an event (pulling out of) that changes its location relative to the background entity (the driveway). But how about (16)? For all we know, the car may keep zooming along for many hours to come. It would be preposterous to claim that we keep our mental simulations zooming along with it for a similar duration. The notion of attentional frame provides an elegant solution to this problem. The situation conveyed by (16) can be conceptualized as the event of perceiving a car zooming along (either from inside or outside of the car). A perceptual event such as this would only take a short amount of time. Along similar lines, a static description like (17) can simply be understood as the event of seeing a blue car rather than as a mental tour de force in which the color blue is continuously being simulated.<sup>4</sup>

In an experiential representation, the perspective of the observer or agent vis à vis the described situation needs to be represented. This is illustrated in (18).

- (18) The red squirrel jumped into the tree.

<sup>4</sup> Often, a speaker or writer will insert a sentence later in the narrative to remind the listener/reader that the car is still in motion. When the reader's attention is directed back to the car, it should still be represented as zooming along.

It is easy to see that the linguistic constructs in (18) by themselves do not provide sufficient perspectival constraints. For instance, is the squirrel jumping from left to right or from right to left? Our working assumption is that in case of underspecification, people will use default expectations, which may depend on environmental constraints, handedness, the direction of reading, or other cultural conventions.<sup>5</sup>

In many cases, however, perspective is automatically produced as part of the mental simulation. Activating relevant perceptual memory traces of experience with words and referents creates a simulation. As a result, the simulation will automatically adopt the perspective of the most frequent (or perhaps most recent) relevant memory traces. For example, clouds are most often seen from below. Therefore, most traces of seeing clouds will be from below. As a result, comprehenders of the sentence: “The farmer looked at the clouds” are likely to simulate clouds from below. In many cases, entity features and the range of the human sensory apparatus jointly place constraints upon the observer’s distance from it. For example, human hearing is such that a whisper can only be heard from a relatively short distance. That is why a whisper in (10) sounds odd. Another illustration is (19), a sentence used by Morrow and Clark (1988).

(19) A mouse/tractor approached the fence.

This example not only shows that the interpretation of “approach” depends on the size of the focal entity—i.e., people place the mouse closer to the fence than they do the tractor (Morrow & Clark, 1988)—it also suggests that the observer is closer to the scene in the

---

<sup>5</sup> For example, right-handers recognize objects more quickly when lit from top-left than when lit from other angles. Presumably this has to do with the fact that under this angle, their right hand does not cast a shadow over manipulated objects. Lefthanders also have a left bias, though less strong than right-handers, presumably because they live in a world dominated by right-handers (Sun & Perona, 1998, but see Mamassian & Goutcher, 2001). Consistent with this left bias, pictures with a left-to-right directionality are judged as more aesthetically pleasing than pictures with a right-to-left directionality (Christman & Pinger, 1997). Recent evidence suggests that bias may be culturally determined (Maass & Russo, in press).

case of the mouse than in the case of the tractor. The main constraint here is the mouse's size, which makes that we can see mice only from relatively short distances. Importantly, this constraint is not just imposed by the mouse's size, but also by the limits of human vision. Hawks, for instance, are able to see mice from much longer distances than humans. In other words, our auditory traces of whispers and our visual traces of mice and tractors already contain relevant perspectival information grounded in human sensation and perception. It may therefore not be farfetched to say that perspective is part of the meaning of some words (Talmy, 2000a, 2000b; Zwaan, in press). There already is some evidence that comprehenders interpret perspective in verbs such as "come" and "go" and "bring" and "take" (Black, Bower, & Turner, 1979). For example, "come" implies movement toward the observer while "go" implies movement away from the observer. In some cases, perspective is explicitly stated. This is the case in a sentence like (20), where the putative default perspective of "cloud" must be overridden.

(20) From the mountaintop, the clouds below looked like big balls of cotton.

To summarize, rather than simply constructing mental representations of who-did-what-to-whom out of mental Lego blocks, we perform experiential simulations that necessarily imply a spatio-temporal perspective on the described situation. The typical experiential perspective on the entity denoted by a linguistic construction plays a key role in establishing a perspective.

As noted earlier, experiential representations consist of multitudes of traces and are stored by the brain in what Damasio (1999) calls 'dispositional form.' Usually, they will only be partly relevant to the current context. Of particular importance therefore is the idea that traces activated by different words constrain each other during construal. For

example, the feature “red” of the focal entity “squirrel” in (18, reprinted below) initially activates a range of traces of visual experiences of the color red.

(21) The red squirrel jumped into the tree.

Most of these traces turn out not to be relevant in the context of the sentence, but that does not become clear until the first noun is processed<sup>6</sup>. Red squirrels are a particular kind of red (brownish red rather than fire truck red). So the noun will constrain what traces are relevant in the current context. But this is only the beginning. Just as red squirrels are a particular kind of red, they are a particular kind of squirrel. Unlike the gray squirrels typically found in North America, they have ear tufts (instead of mouse-like ears) and are smaller. In other words, the two concepts, “red” and “squirrel,” constrain each other’s representation. One way of conceptualizing this is that all the traces that make up a dispositional representation receive some degree of activation from the associated word, but that only one or a few of them will be activated above threshold and thus become incorporated in the mental simulation. In the next cycle, “jumped” provides further constraints on the ongoing simulation. For one, it provides some articulation of the shape of the squirrel; it is stretched out, rather than sitting on its hind legs, for example.<sup>7</sup>

During construal, the information is being integrated with previous construals, which form part of the context for the current construal in the comprehension of connected discourse. This is where the remaining two components of construal come into play: time frame and spatial region. When two construals pertain to the same time

---

<sup>6</sup> As evidence shows, comprehension is incremental, rather than postponed until certain linguistic boundaries are reached (e.g., Chambers et al., 2001).

<sup>7</sup> Also note that ‘squirrel’ constrains the interpretation of ‘jumped.’ After all, the way a squirrel jumps is different from the way a human or an antelope jumps.

interval, they can be integrated more easily than when they pertain to two different intervals (Zwaan, 1996). To a certain extent, this is also true for spatial regions (see Zwaan & Radvansky, 1998 for a discussion). The topic of integration is beyond the scope of the present chapter, but is discussed in Zwaan (in press).

#### 4. Empirical evidence

What empirical evidence do we have for our claims? In this section we review the research from our lab; further relevant evidence is reviewed in other chapters in this volume. In most experiments, we used the same methodology. Subjects are exposed to linguistic materials and are then presented with one or more pictures. Their task consists in comprehending the sentences and performing speeded judgments on the pictures. The rationale is that sentence comprehension will involve a construal in which visual traces are activated and used that either match or mismatch the visual traces created by the pictures. 'Match' and 'mismatch' should be thought of in relative terms only. In our interactions with the environment there will probably never be a perfect match between a new visual trace and one already in memory. For example, we rarely if ever see objects under identical angles and lighting conditions and against identical backgrounds on different occasions. As a result, all visual traces are slightly different from each other. Some theories of object recognition account for this by assuming that some amount of interpolation between visual traces in memory occurs in order to obtain a match with the visual input (e.g., Bühlhoff & Edelman, 1992; Tarr, 1995). In the context of our experiments, the claim is simply that the match condition provides a stronger match

between construal and picture traces than the mismatch condition. We should also note that in our experiments, subjects are often not directly relating the picture to the sentence. In that sense, it cannot be argued that the subjects' responses are due to special imagery strategies.

Here is a list of claims we have made about construal:

1. Comprehenders represent perceptual aspects of referents or situations;
2. Comprehenders represent spatial relations between object parts;
3. Comprehenders represent dynamic aspects of events;
4. Comprehenders represent perspective.

How do these claims hold up against empirical scrutiny? Let's revisit sentences (3) and (4), which are included below as (22) and (23).

(22) John pounded the nail into the wall.

(23) John pounded the nail into the floor.

Our contention is that these sentences lead to different mental representations.

Specifically, we predict that comprehenders will represent the orientation of the nail. In order to test this prediction, Stanfield and Zwaan (2001) presented subjects with sentences such as (22) and (23), followed by a line drawing of an object. The subjects simply decided whether the picture depicted a word mentioned in the sentence. Two things are important to note. First, a picture of a nail, irrespective of its orientation, should yield a "yes" response. Second, a picture of a horizontal nail would match a construal of (22) and one of a vertical nail would match a construal of (23). Stanfield and Zwaan found that responses were significantly faster in the match than in the mismatch condition. One sub-optimal feature of these experiments was that the direction in which

the nail points is indeterminate in case of the wall— it could point to the left or to the right (for obvious reasons, we don't have this problem with the floor). As mentioned earlier, it may be that people use a default assumption in cases such as these. But this means that the direction of the nail would in some cases mismatch that of the nail in the construal. Our counterargument is that the match condition still provides a better match than the mismatch condition because the visual trace of a horizontal nail, regardless of its orientation, provides a better match than a picture of a vertical nail.<sup>8</sup>

In a later series of experiments, we tested the claim that construal necessarily includes a representation of the shape of a focal entity (Zwaan, Stanfield, & Yaxley, 2002). Using the same logic as in our earlier experiments, we had subjects read sentences and then presented them with pictures. The sentences were of the type of (9) and (10) or as (24) and (25).

(24) He saw the lemon in the bowl.

(25) He saw the lemon in the glass.

In addition to a recognition task, we also employed a naming task. In a naming task, the subject sees a picture and simply names it. We used a naming task because it provides a more implicit measure than a recognition task. The recognition task calls for the subject to compare the picture with the sentence. Not so in the naming task—naming the picture does not directly involve reference to the sentence. Nonetheless, in both experiments we found that the match condition yielded faster responses than the mismatch condition.

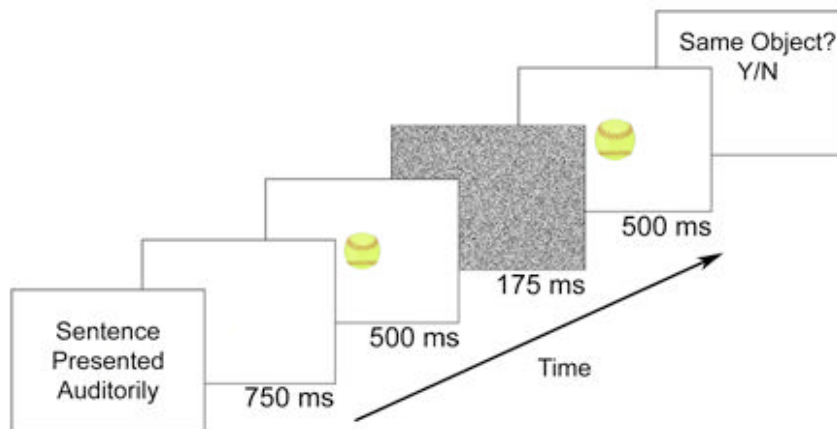
These findings suggest that comprehenders routinely represent the shape of the focal entity mentioned in a sentence.

<sup>8</sup> MacWhinney (this volume) points out that the orientation of the nail is only part of the mental simulation, not the whole simulation. We concur. It is, however, a diagnostic part of the simulation and as such is amenable to an empirical test.

In a more recent experiment, we investigated whether comprehenders form perceptual representations of described motion (Zwaan, Madden, Yaxley, & Aveyard, submitted). Subjects listened to sentences such as (26) or (27) over headphones.

(26) The shortstop hurled the softball at you.

(27) You hurled the softball at the shortstop.



*Figure 3. Schematic depiction of an experimental trial in Zwaan et al. (submitted).*

After each sentence, they saw two pictures each presented briefly and separated by a mask (see figure above). On critical trials, the depicted object was mentioned in the sentence (e.g., a softball). Crucially, the second picture was either bigger or smaller than the first one, thus suggesting movement toward or away from the viewer. The size changes were very subtle. Subjects judged whether the two pictures were the same. On trials requiring “no” responses, the two pictures were of different objects (e.g., a basketball and a snowmobile). In other words, the picture-judgment task was extremely easy. Nonetheless, the subjects’ responses were influenced by the content of the sentences, exactly as predicted by our construal theory. Picture sequences in which the

second ball was bigger than the first were judged significantly faster when the sentence implied movement toward the protagonist (e.g., as in (26)) than when the sentence implied movement away from the protagonist (as in (27)). And the reverse was true for sequences in which the second ball was smaller than the first. Given that the sentence content was irrelevant to the picture-judgment task, these results support the idea that comprehenders spontaneously represent motion.

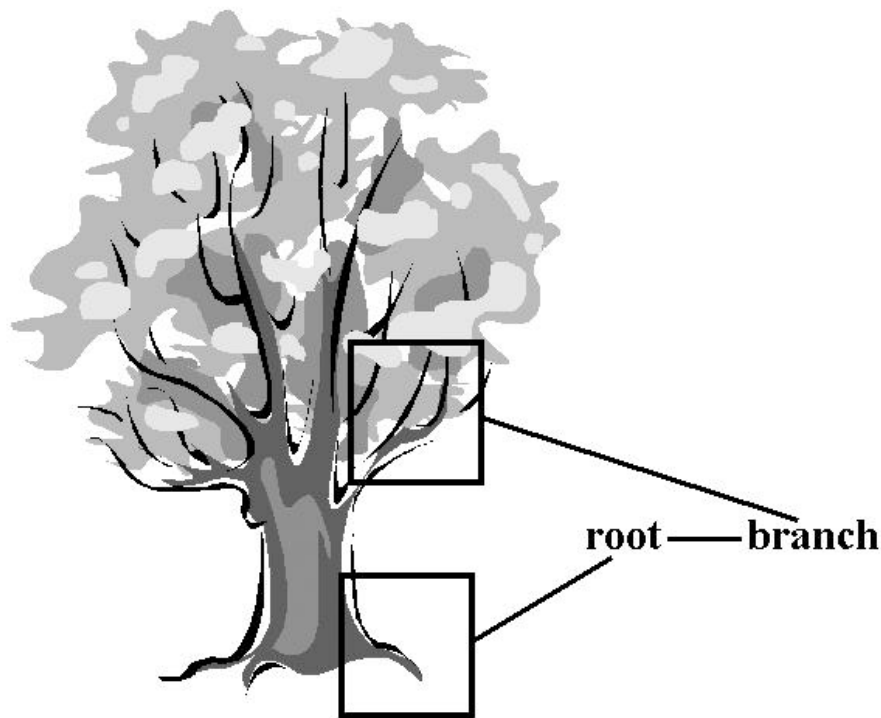
We explain this pattern by assuming that over the course of their lives, people have stored numerous traces in memory of objects moving toward them—and occupying an increasingly larger segment of their visual field— or away from them—and occupying an increasingly smaller segment of their visual field. These traces are dynamic representations (Freyd, 1987; Wallis & Bühlhoff, 1999) in that they extend and change over time. In accordance with Wallis and Bühlhoff, among others, we assume that dynamic mental object representations are the result of spatiotemporal associations between visual patterns acquired during experience of our environment. The picture sequence activates these traces as does construal of the preceding sentences. As a consequence, in the match condition the relevant visual traces are already activated via construal before the picture sequence is seen. In the mismatch condition, a dynamic trace is activated by the linguistic stimuli that is the reverse of the one activated by the picture sequence.

Another claim we made is that comprehenders represent the spatial relations between referents. We found evidence for this using a paradigm that was slightly different from the paradigms discussed before. In this paradigm, subjects made speeded semantic-related judgments to word pairs, e.g., branch—root. The key feature of the

manipulation was in the locations of the words on the computer screen. In the match condition, the words were presented in a manner that was consistent with the relative positions of their referents. For instance, branches are canonically above roots and so the word branch was presented above root in the match condition. In the mismatch condition, the positions of the words were reversed, so that the word denoting the top referent was now presented below the word denoting the bottom referent. If people simply use lexical associations to make these judgments, the relative positions of the words on the screen should not make a difference. However, if people rely on perceptual representations to make these judgments, the match condition should yield faster responses than the mismatch condition. And this was exactly what we found (Zwaan & Yaxley, in press; Experiments 1 and 3). Importantly, the effect disappeared when the words were presented horizontally, ruling out that it was caused by the order in which the words were read (Experiments 2 and 3). In a later study, using a visual-field manipulation, we found that the mismatch effect only occurred when the word pairs were briefly flashed to the left visual field and thus processed by the right hemisphere (Zwaan & Yaxley, 2003). This is consistent with the well-known fact that the right hemisphere is more involved in processing spatial relations than the left hemisphere.

Our account for these findings is straightforward. Parts of objects, such as branches, wheels, or elbows are typically not seen in isolation. In fact, they derive their meaning from context. For example, by itself “elbow” means nothing. It derives its meaning from being part of an arm. In perceptual terms it can be conceptualized as the focal part of an attended object (Langacker, 2001). When a word like “branch” is read or heard, visual traces of trees with branches as the focal part will be activated. And when

“root” is read or heard, visual traces of trees with roots as the focal part will be activated. The spatial positions of the referents relative to the larger entities (landmarks) are part of the activated representation. The positioning of the words on the screen produces its own visual trace. This is either consistent or inconsistent with the composite visual traces of the referents thus producing the mismatch effect (see Figure 5).



*Figure 5. The relations between perceptual traces of roots and branches and between the lexical constructions associated with them.*

We are not claiming that the semantic-relatedness judgments are exclusively based on visual traces of the words' referents. As pointed out in the introduction and as

shown in Figure 1, experiential traces of referents are connected to experiential traces of linguistic constructs, which are interconnected based on co-occurrence in the linguistic input and output streams. Semantic-relatedness judgments are presumably based both on lexical and on referent associations. The left hemisphere data in Zwaan and Yaxley (2003) support this idea. The relative positions of the words on the screen did not affect the subjects' judgments, but these judgments were still accurate (i.e., the words were seen as related), implicating lexical associations.

We have reviewed several experiments from our lab that collectively have produced evidence in support of the view that construals involve visual representations. However, it is important to note that visual aspects are only one type—albeit a very important one—of experiential representations. Research currently underway in our lab focuses on auditory information as well.

A central claim of our theory concerns perspective. As yet, we have not conducted any specific studies on this topic, although the experiments on dynamic mental representations and on resolution are broadly relevant. The former demonstrate that the comprehenders had taken the perspective of the protagonist, given how their responses to picture sequences implying movement of an object toward or away from them were influenced by their correspondence with the movement described in the sentence, i.e., toward or away from the protagonist. We clearly need to have more extensive research on whether and how perspective is established during sentence comprehension. One important approach is described by MacWhinney (this volume).

## 5. Abstract representations

The empirical evidence we have discussed so far pertains to concrete situations. A common criticism of embodied theories is that they are ill equipped to deal with abstract information. However, several approaches to this problem have been advanced. One such approach holds that abstract representations are created from concrete representations by way of (metaphorical) extension (Gibbs, this volume; Goldberg, 1995; Lakoff, 1987). Indeed, there is empirical evidence for this view (Boroditsky & Ramscar, 2002). Langacker (1987) has likened the abstraction process to superimposing multiple transparencies. What is not common across transparencies will become blurred, but commonalities will become well defined. For example, children often hear expressions such as “Put the ball on the floor”, “Put the cup on the table” The only thing that is constant across all instances is that the child uses his or her hands to move an object to a location and then release it. This commonality then becomes schematized in action and language as “Put X” (Tomasello, 2003). Talmy (1996) assumes that a similar abstraction process underlies the meaning of locative prepositions such as ‘across.’<sup>9</sup>

At first sight, a notion like negation presents a major problem for embodied accounts of meaning. How does one mentally simulate an entity, event or feature that is not present? Cognitive linguists have proposed to view negation as a sequence of simulations. We first simulate the situation that is negated and then the actual situation (Fauconnier, 1985). Thus, the meaning of “not” is captured by a sequence of construals rather than by a dictionary definition. We tested the first part of the negation hypothesis—that people initially represent the negated situation— by modifying the

---

<sup>9</sup> The point here is not that the embodied account of abstraction is unique. Rather, the point is that there exist embodied accounts of abstraction.

sentences from Zwaan et al. (2002). We simply converted the affirmative sentences into negative sentences, such as (30).

(30) The eagle was not in the tree.

If comprehenders first construe the negated situation, then they should initially show the same mismatch effect as the subjects in Zwaan et al. (2002). For example, (30) should first give rise to a construal of an eagle in a tree (wings drawn in). We did indeed obtain this effect in three experiments (Kaup, Yaxley, Madden, & Zwaan, in preparation).

Obviously, the same mismatch effect would have occurred if our subjects had simply ignored the negation operator. However, we included comprehension questions (e.g., “Was the eagle in the nest?”), which prompted the subjects to pay attention to the negation. Most subjects answered these questions accurately. Moreover, the mismatch effect was still present when we analyzed the data from subjects with accuracy greater than 90% only. In other words, we can be confident that the subjects did indeed process the negation.

These findings demonstrate that comprehenders first construe the negated situation. But this is only part of the story, of course. It is now important to demonstrate that they will end up construing the actual situation. This will probably require the use of different items, as the actual situation is not strongly constrained by our materials. For instance, if the eagle is not in the nest, it doesn't mean that it is in the sky. It could be in a tree or perched on a rock, in both of which cases it would not have its wings outstretched. Nevertheless, this initial foray into finding empirical evidence for an experiential account of a rather abstract concept like negation is promising (see Kaup, Lüdtké, & Zwaan, in press, for a more thorough discussion). Other approaches to an embodied account of

abstractions are being discussed in other chapters in this book (Barsalou & Wiemer-Hastings, this volume; Gibbs, this volume; Prinz, this volume).

## 6.0 Conclusion and outlook

We started this chapter by comparing two perspectives on cognition in general and on language comprehension in particular. According to the traditional perspective, language comprehension involves the activation and integration of discrete and abstract building blocks of meaning. These building blocks are most commonly represented as propositions or equivalently in semantic networks (in which a pair of nodes are the arguments and the link between them the predicate). We have demonstrated that this general view cannot account for a range of recent findings about language comprehension obtained in our experiments and in experiments by others. According to the alternative view, cognition in general and language comprehension in particular involve the activation and integration of experiential traces in the construal of a situation. These traces are activated by linguistic constructs, which are experiential representations in their own right. Language can be viewed as a sequence of cues modulating the comprehender's attention to a referential world, which is simulated by integrating experiential traces. As we have shown, this view can account for the recent findings we have discussed in this chapter. Moreover, it generates a volley of testable predictions regarding language comprehension, for example about the role of perspective in language comprehension.

However, it is clear that the view we have attempted to outline here is in need of further articulation. The challenge for researchers adopting an experiential perspective is to further articulate the theoretical frameworks, keeping them consistent with what is known about the brain, and test them in elegant and convincing experiments. We are optimistic that these goals are within our reach.

Our focus has been on the role of visual representations in language comprehension. This is because our empirical research thus far has focused on this phenomenon, primarily because it was motivated in part by the goal to show the limitations of amodal propositional representations. However, it is clear that embodied language comprehension involves more than just visual representations. For example, there is behavioral evidence that language comprehension may involve the activation of motor programs (Glenberg & Kaschak, 2002). The sympathetic activation of neurons in the premotor cortex, first observed in monkeys, is thought to underlie action understanding and therefore to mediate language comprehension (Rizzolatti & Arbib, 1998). MacWhinney (this volume) provides an extension of these ideas. A full-fledged theory of embodied language comprehension should include both perceptual and action simulations. Moreover, it should be capable of explaining their relations (do they occur simultaneously, are they integrated into a coherent representation or are they independent, how is their construal cued by linguistic constructs).

The claims made by proponents of embodied comprehension, for example about the activation of visual representations during language comprehension, may at the same time seem trivial and counterintuitive. They will seem trivial to the lay person, or even to people with great expertise in the use of language, such as novelists and poets. Of course,

words can be used to conjure up images in the reader's mind! However, these same claims will seem counterintuitive to researchers trained in traditional cognitive science. To them, the claim that meaning can be captured by experiential representations does not make sense. For one, the claim opens the door to the homunculus problem, and thus to an infinite regress. If there are pictures in the head, then there must be a little person in there looking at the pictures. And if so, who's in that person's mind? There are two responses to this criticism. First, this problem also seems to apply to the amodal view. After all, where is the little person reading all those quasi-linguistic propositions entering and leaving the revolving door of working memory? Second, and more importantly, the claim is not that there are pictures in the mind. Rather, the claim is that traces of visual and other experiences are (partly) reactivated and recombined in novel ways by associated words (Barsalou, 1999). In this sense, language comprehension is the vicarious experiencing of events. Speakers and writers carefully orchestrate linguistic constructs so that experiential traces in their audience's minds can be recombined in novel ways to produce novel experiences.

Author note.

We thank Diane Pecher and two anonymous reviewers for helpful feedback on an earlier version of this manuscript. The research reported here was supported by grant MH-63972 from the National Institutes of Health. The chapter was written while the first author was a Fellow at the Hanse Institute for Advanced Study in Delmenhorst, Germany. Please address all correspondence regarding this paper to: Rolf A. Zwaan, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270. Email may be sent to: [zwaan@psy.fsu.edu](mailto:zwaan@psy.fsu.edu).

References

- Barsalou, L.W. (1999). Perceptual Symbol Systems. Behavioral and Brain Sciences, *22*, 577-660.
- Black, J. B., Turner, E., & Bower, G. H. (1979) Point of view in narrative comprehension memory. Journal of Verbal Learning and Verbal Behavior, *18*, 187-198.
- Bloom, L. (2000). Pushing the limits on theories of word learning. Monographs of the Society for Research in Child Development, *65*, 124-135.
- Boroditsky, L. & Ramscar, M. (2002). The roles of body and mind in abstract thought. Psychological Science, *13*, 185-188.
- Bülthoff, H. & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Academy of Sciences, USA *89*, 60-64.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. Papers and Reports on Child Language Development, *15*, 17-29.
- Chafe, W. (1994). Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing. Chicago, IL: University of Chicago Press.
- Chambers, C.G., Tanenhaus, M.K., Eberhard, K.M., Filip, H., & Carlson, G.N. (2001). Circumscribing referential domains in real-time language comprehension. Journal of Memory and Language, *47*, 30-49.

- Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton.
- Christman, S., & Pinger, K. (1997). Lateral biases in pictorial preferences: pictorial dimensions and neural mechanisms. Laterality, *2*, 155-175.
- Damasio, A.R., (1999). The feeling of what happens: body and emotion in the making of consciousness. Harcourt Brace & Company.
- Edelman, S. (2003). Naturalizing Linguistics. Manuscript submitted for publication.
- Fauconnier, G. (1985). Mental spaces: Aspects of meaning construction in natural language. Cambridge: MIT Press.
- Fodor, J.A. (1983) The Modularity of Mind: An Essay on Faculty Psychology Cambridge, MA: MIT Press.
- Freyd, J.J. (1987) Dynamic mental representations. Psychological Review, *94*, 427-438.
- Givón, T. (1992). The grammar of referential coherence as mental processing instructions. Linguistics *30*, 5-55.
- Glenberg, A.M. (1997). What memory is for. Behavioral and Brain Sciences *20*, 1-55.
- Goldberg, A.E., (1995). Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.
- Goldberg, A.E., (2003). Constructions: A new theoretical approach to language. Trends in Cognitive Science, *7*, 219-224.
- Harnad, S. (1990) The Symbol Grounding Problem. Physica D, *42*, 335-346.

- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.
- Hockett, C.F. (1959). Animal 'languages' and human language. Human Biology 31, 32-39.
- Kaup, B., Zwaan, R.A., & Lüdtke, J. (in press). The experiential view of language comprehension: How is Negation Represented? In: F. Schmalhofer & C.A. Perfetti (Eds.) Higher language processes in the brain. Mahwah, NJ: Erlbaum.
- Kaup, B., Yaxley, R.H., Madden, C.J., & Zwaan, R.A. (in preparation). Perceptual Simulation of Negated Text Information.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.
- Lakoff, G. (1987) Women, fire, and dangerous things: what categories reveal about the mind. Chicago: University of Chicago Press.
- Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.
- Langacker, R. (1987). Foundations of cognitive grammar, Vol. 1, Stanford, CA: Stanford University Press.
- Langacker, R.W. (2001). Discourse in cognitive grammar. Cognitive Linguistics 12, 143-188.
- Maass, A., & Russo, A. (in press). Directional bias in the mental representation of spatial events: nature or culture? Psychological Science.

- Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. Cognition, 81, B1-B9.
- Morrow, D. G., & Clark, H. H. (1988). Interpreting words in spatial descriptions. Language and Cognitive Processes, 3, 275-291.
- Pustejovsky, J. (1995). The generative lexicon. Cambridge, MA: MIT Press.
- Pylyshyn, Z.W. (1986). Computation and cognition: Toward a foundation for cognitive science. Cambridge, MA: MIT Press.
- Sadoski, M., & Paivio, A. (2001). Imagery and text: A dual coding theory of reading and writing. Mahwah, NJ: Erlbaum.
- Searle, J.R. (1980). Minds, brains, and programs. Behavioral & Brain Sciences, 3, 417-457.
- Stanfield, R.A. & Zwaan, R.A. (2001). The effect of implied orientation derived from verbal context on picture recognition. Psychological Science, 12, 153-156.
- Sun, J., & Perona, P. (1998). Where is the sun? Nature Neuroscience, 1, 183-184.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18, 645-660.
- Talmy, L. (1996). Fictive motion in language and 'ception.' In: Bloom, P., Nadel, L., & Peterson, M.A. (Eds.). Language and space (pp. 211-276). Cambridge, MA: MIT Press.
- Talmy, L. 2000a. Toward a Cognitive Semantics, vol. 1, Concept Structuring Systems. Cambridge, MA: MIT Press

Talmy, L. 2000b. Toward a Cognitive Semantics, vol. 2, Typology and Process in Concept Structuring. Cambridge, MA: MIT Press.

Tarr, M. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psychonomic Bulletin & Review, 2, 55-82.

Ter Meulen, A. G. B. (1995). Representing time in natural language: the dynamic interpretation of tense and aspect. Cambridge, MA: MIT Press.

Tomasello, M. (2003). Constructing a language. A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Vendler, Z. (1967). Linguistics in philosophy. Ithaca, NY: Cornell University Press.

Wallis, G., & Bühlhoff, H. (1999). Learning to recognize objects. Trends in Cognitive Sciences, 3, 22-31.

Zwaan, R. A. (1996). Processing narrative time shifts. Journal of Experimental Psychology: Learning, Memory, and Cognition 22, 1196-1207.

Zwaan, R.A. (in press). The immersed experiencer: toward an embodied theory of language comprehension. In: B.H. Ross (Ed.), The Psychology of Learning and Motivation, Vol. 44. New York: Academic Press.

Zwaan, R.A., Madden, C.J., Yaxley, R.H., & Aveyard, M. (submitted). Moving words: Language comprehension produces representational motion.

Zwaan, R. A., & Radvansky, G. A. (1998) Situation models in language comprehension and memory. Psychological Bulletin, 123, 162-185.

Zwaan, R.A., Stanfield, R.A., Yaxley, R.H. (2002). Do language comprehenders routinely represent the shapes of objects? Psychological Science, *13*, 168-171.

Zwaan, R.A., & Yaxley, R.H. (in press). Spatial iconicity affects semantic-relatedness judgments. Psychonomic Bulletin & Review.

Zwaan, R.A., & Yaxley, R.H. (2003). Hemispheric differences in semantic-relatedness judgments. Cognition, *87*, B79-B86.

Author note

Rolf A. Zwaan and Carol J. Madden, Psychology Department, Florida State University, Tallahassee, FL 32306-1270. Correspondence regarding this chapter should be addressed to the first author at [zwaan@psy.fsu.edu](mailto:zwaan@psy.fsu.edu). The research reported in this chapter is supported by grant MH-63972 to R.A.Z. This chapter was written in part while the first author was a Fellow at the Hanse Institute for Advanced Study in Delmenhorst, Germany.