

THE IMMERSED EXPERIENCER:  
TOWARD AN EMBODIED THEORY OF LANGUAGE COMPREHENSION

Rolf A. Zwaan

Florida State University

KEY WORDS: comprehension, embodied cognition, language, mental representation,  
mental simulation, situation model

To appear in B.H. Ross (Ed.) The Psychology of Learning and Motivation, Vol. 44. New  
York: Academic Press.

Address all correspondence to:  
Dr. 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

Table of Contents

- I. Introduction
- II. Comparison with other frameworks
- III. Components of the comprehension process
- IV. Activation
- V. Construal
  - A. Definition of construal
  - B. Components of construal
  - C. Summary
- VI. Integration
  - A. Definition of integration
  - B. Types of transitions
  - C. Factors influencing integration
- VII. Empirical evidence consistent with the IEF
  - A. Words activate of perception/action-related brain areas
  - B. Action representations are activated during comprehension
  - C. Perceptual representations are routinely activated during comprehension
  - D. Comprehenders respond as if “in” the narrated situation
- VIII. Accounting for propositional findings
- IX. Accounting for abstract language
- X. Conclusions

## I. Introduction

Consider the following four sets of findings about language comprehension.

- (1) Words activate brain regions that are close to or overlap with brain areas that are active during perception of or actions involving the words' referents (Isenberg et al., 2000; Martin & Chao, 2001; Pulvermüller, 1999, 2002). Brain lesions in patients with selective semantic impairments affect perceptual representations (e.g., Farah & McClelland, 1991; McRae, de Sa, & Seidenberg, 1997; Miceli et al., 2001).
- (2) Visual representations of object shape and orientation are routinely and immediately activated during word and sentence comprehension (Dahan & Tanenhaus, 2002; Stanfield & Zwaan, 2002; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Yaxley, in press a, b). Visual-spatial information primes sentence processing (Boroditsky, 2000) and may interfere with comprehension (Fincher-Kiefer, 2001).
- (3) Information that is "in" the situation described in a text is more active in the comprehender's mind than information that is not in the situation (Glenberg, Meyer, & Lindem, 1987; Kaup & Zwaan, in press; Morrow, Greenspan, & Bower, 1987; Horton & Rapp, in press; Trabasso & Suh, 1993; Zwaan, Madden, & Whitten, 2000).
- (4) When comprehending language, people's eye and hand movements are consistent with perceiving or acting in the described situation (Glenberg & Kaschak, in press; Klatzky, Pellegrino, McCloskey, & Doherty, 1989; Spivey, Richardson, Tyler, & Young, 2000).

These bodies of findings present a challenge to current theories of language comprehension, and there currently exists no framework that coherently accounts for them. The purpose of this contribution is to develop such a framework. The basic premise is that language is a set of cues to

the comprehender to construct an experiential (perception plus action) simulation of the described situation. In this conceptualization, the comprehender is an immersed experiencer of the described situation and comprehension is the vicarious experience of the described situation. Consider the following examples.

(1a) The ranger saw the eagle in the sky.

(1b) The ranger saw the eagle in the nest.

Amodal propositional representations (e.g., Kintsch & van Dijk, 1978) of these sentences would look as follows.

(2a) [[SAW[RANGER, EAGLE]], [IN[EAGLE, SKY]]]

(2b) [[SAW[RANGER, EAGLE]], [IN[EAGLE, NEST]]]

Although these representations capture the expressed relations among entities, actions, and locations, they leave out something critical about the eagle: its shape. When in the sky, an eagle has its wings outstretched and when in a nest, it has its wings drawn in. The Immersed Experiencer Framework (IEF) proposed here is capable of capturing information such as this. The basic idea is that words activate experiences with their referents. For example, sentence 1a activates visual experiences of eagles in the sky. In these visual representations, the eagle has its wings outstretched. On the other hand, 2b activates visual experiences of seeing eagles in a nest. In these representations, the eagle has its wings drawn in. Thus, the shape of the referent falls out from the perceptual simulation.

The IEF distinguishes three component processes of language comprehension: activation, construal, and integration. Activation operates at the word level, construal at the clause level, and integration at the discourse level. The IEF coherently accounts for the findings alluded to in the opening paragraph. In addition, it can account for the findings that earlier models can account for.

As a consequence, the IEF provides a useful way of conceptualizing language comprehension, although there are significant hurdles to be taken.

## II. Comparison with other frameworks

Fortunately, research on language comprehension is cumulative. The IEF adopts many insights from earlier theories of comprehension, specifically those developed by Kintsch and van Dijk (Ericsson & Kintsch, 1995; Kintsch, 1988; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983), as well as the related event-indexing model (Zwaan & Radvansky, 1998). In addition, there is overlap with other well-known approaches to comprehension, such as the Structure-Building-Framework (Gernsbacher, 1990), and the constructivist framework (Graesser, Singer, & Trabasso, 1994). Along with Kintsch and van Dijk (1978), I assume that (1) the linguistic input stream is segregated into units, which are subsequently integrated with the contents of working memory and (2) that comprehension proceeds in an incremental fashion, whereby currently relevant information (constructed from previous input) is held in an active state so that it influences the integration of incoming information. Along with Ericsson and Kintsch (1995), I assume that parts of long-term memory are recruited for this process along with short-term working memory, thus allowing the comprehender to maintain relevant aspects of multiple construals active in working memory, thus increasing the likelihood of overlap between the current construal and the contents of working memory.

Along with Kintsch (1988), I assume that incoming words first result in a diffuse pattern of activation, which is subsequently narrowed down by a constraint satisfaction mechanism that takes contextual information into account. However, as will become clear, I am assuming that initial activation is even more diffuse than was assumed by Kintsch. Along with many researchers (e.g.,

Bower & Morrow, 1990; Glenberg, Meyer, & Lindem, 1987; Graesser, Singer, & Trabasso, 1994; Johnson-Laird, 1983, Sanford & Garrod, 1981; van Dijk & Kintsch 1983; Zwaan & Radvansky, 1998), I assume that the typical goal of language comprehension is the construction of a mental representation of the referential situation, a situation model. Along with Gernsbacher (1990) and Zwaan and Radvansky (1998), I assume that the on-line comprehension process is strongly influenced by spatio-temporal characteristics of the referential situation, in addition to characteristics of the linguistic input stream.

Although the IEF adopts assumptions and mechanisms from these earlier models, it differs from them in that it makes the explicit assumption that comprehension involves action and perceptual representations and not amodal propositions. This view is informed and inspired by theories that ground cognition in perception and action (e.g., Barsalou, 1999; Damasio, 1994, 1998; Glenberg, 1997; Lakoff, 1987; Langacker, 1987; MacWhinney, 1999; Sadoski & Paivio, 2001; Pulvermüller, 1999, 2002; Talmy, 1988). For instance, Pulvermüller (1999; 2002) has proposed a Hebbian model of semantic representation according to which the perception of a word activates “functional webs” of neurons located throughout the cortex. These functional webs are also activated when the word’s referent is experienced. As such, comprehension of a word is the reconstitution of an experience with its referent. It is too rigid to assume that words will activate a single web. Which parts of the functional web will be activated depends on the semantic and task context in which the word is processed (Posner & DiGirolamo, 1999). This context-sensitivity is achieved in the IEF, as will become clear later.

In short, the basic assumption is reading or hearing a word activates experiential representations of words (lexical, grammatical, phonological, motoric, tactile) as well as associated experiential representations of their referents –motor, perceptual, and emotional representations,

and often combinations of these (see also Sadoski & Paivio, 2001). These traces can be activated by verbal input and as such enable the reconstitution of experience. In this sense, then, comprehension is the vicarious experience of the described events through the integration and sequencing of traces from actual experience cued by the linguistic input (see Duchan, Bruder, & Hewitt, 1995, and Gerrig, 1993 for other experiential views).

### III Components of the comprehension process

INSERT TABLE 1 ABOUT HERE

The IEF distinguishes three general components of the comprehension process: activation, construal, and integration. Table 1 lists these processes, along with the linguistic and representational units on which they operate, and the referential units that they denote. It is important to stress that the three component processes are not assumed to operate sequentially. Rather, it is assumed that they have a large degree of temporal overlap. However, for reasons of exposition, they will be described sequentially. Table 1 also shows the linguistic units that constitute the input for these processes. It is assumed that words are associated with a range of experiential traces related to their referents, functional webs. Finally, the fourth column lists the referential units, that is, the entities or events in the environment that are associated with the representational units.

### IV. Activation

Incoming words activate functional webs that are also activated when the referent is experienced. Thus, in the case of sentences 1a and 1b, the functional webs that are also active when we see an eagle will be activated. Functional webs are widespread throughout the cortex and may

involve the primary sensory areas (Pulvermüller, 2002). Given that we usually have different experiences with referents (e.g., we see objects from different viewpoints), there will initially be diffuse activation of multiple overlapping functional webs. For example, a functional web encoding experiences of seeing eagles in flight (from different perspectives) may be activated, as well as a functional web encoding visual experiences of perched eagles.

Thus, assuming there is no prior semantic context (although this is rarely the case outside the cognition lab and a game of Scrabble) and our word is the first content word of the sentence being processed, this word will diffusely activate overlapping functional webs. This web comprises the totality of our experiences with a certain entity or event. The degree of diffuseness of the representation depends on the frequency distributions and primacy and recency of our experiences with its referent. If a certain category of experiences, for instance a specific visual perspective, is relatively frequent for that entity compared to other experiences with the same entity (e.g., we most often see hot air balloons from below, rather than from above and we most often see eagles from afar rather than from up close) then the most frequent experiential trace will be the most highly activated (see Palmer et al., 1981 on canonical viewpoints). When there are experiential categories with roughly similar frequencies, then the initially activated functional web will not be strongly biased toward a specific representation. The activated representation will provide the context for the pattern of activation for the next functional web. This provides a constraint on the activation of the new web. The more specific, or “articulated,” the initial representation, the stronger the constraints it exerts on the subsequent activation. The constraint-satisfaction mechanism by which a previously diffusely activated functional web is constrained to fit the mental simulation is called articulation, and it occurs during construal.

## V Construal

### V.A. Definition of construal

Construal is the integration of functional webs in a mental simulation of a specific event. The grammatical unit on which construals operate is an intonation unit. This is grounded in the analysis of spoken language, which precedes written language both phylogenetically and ontogenetically. Chafe (1994, p. 69) has observed that speech can be segmented into units, called intonation units, that are identifiable on the basis of cues such as pauses, pitch shifts, and changes in voice quality. They often coincide with clauses, but not necessarily so. For example, a single clause sentence like (3) contains three intonation units.

(3) We will meet in my office on the fourth floor of the Longmire building.

The location of the meeting is specified in three intonation units, which convey a zooming-out from the office to the floor it is on to the building it is in. In terms of comprehension, intonation units can be viewed as attentional frames (Langacker, 2001). This notion is particularly insightful, because it construes language comprehension as the language-based modulation of attention to a described state of affairs.

During construal, initially activated functional webs are integrated to yield a representation of an event. The referential unit to which construal pertains is an event. During construal, the initially diffusely activated functional webs become articulated by way of a constraint-satisfaction mechanism. The general principles of Kintsch's (1988) construction-integration model, not developed with perceptual symbol systems in mind, provide a way of conceptualizing this process. In the case of sentences 1a and 1b, the constraint is provided by the prepositional phrase, which states the location of the eagle (in the sky vs. in the nest). Simplifying a bit for the sake of argument, the prepositional phrase constrains the activated visual representations (functional webs),

such that those consistent with the stated location (i.e., our visual representations of eagles in the sky or eagles in a nest) will receive more activation than those inconsistent with the stated location. As a result, the appropriate visual representation of the eagle in its location will be the end result of construal. The visual representation of the eagle, which initially was diffuse because multiple shapes were activated is now articulated. Given that language comprehension normally occurs rapidly—normal speech rate is about 2.5 words per second and normal reading rate is about twice as fast--and is incremental--the comprehender attempts to interpret each word immediately (Chambers, Tanenhaus, Eberhard, Filip, & Calson, 2001), the articulation of a functional web in a mental simulation occurs rapidly as well.

Methods with high temporal resolution (because they do not involve manual or vocal responses), such as eye tracking and event-related potentials have provided substantial evidence that the incorporation of incoming information occurs immediately rather than at some syntactic boundary and incorporates not only clausal information, but also information from previous clauses (e.g., Chambers et al., 2001; van Berkum, Brown, & Hagoort, 1999). Thus, construal is an immediate and incremental process. In some cases, the functional web activated by one word may constrain the activation of the functional web activated by the next word (forward articulation), while the latter may at the same time constrain the former (backward articulation). This process of mutual articulation can be illustrated by the following example (see also Morrow and Clark, 1988; and Half, Ortony, & Anderson, 1976). Consider (4a) and (4b).

(4a) The red squirrel jumped from the oak to the pine tree.

(4b) The red fire truck came swerving around the corner.

In both sentences, “red” will activate visual representations of the range of colors we call “red.”

The next content word helps articulate a specific instance of red, a brownish red in the case of the

squirrel and a bright red in the case of the fire truck (the more prototypical red). In turn, the first word also plays a role in articulating the symbol activated by the second word. As a consequence, we now have a representation of a red squirrel—a smallish squirrel with ear tufts, common in Europe—rather one than of a gray squirrel—a slightly bigger squirrel with mouse-like ears, common in North America. Obviously, this only holds for comprehenders with the requisite visual experiences. It is important to note that many aspects of the squirrel remain unarticulated, such as the color of its eyes, the bushiness of its tail, and so on. Construals are necessarily schematic. This is so for three reasons. First, language profiles a situation in a certain way, directing attention to some aspects of the situation, but not others (Langacker, 1987). Secondly, perception itself is limited by attentional capacity and thus the experiential traces that are used during comprehension are schematic to begin with (Barsalou, 1999; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). Third, comprehenders use an economy of processing (especially in the psychological laboratory), and typically do not activate more information than is necessary for comprehension (Graesser, Singer, & Trabasso, 1994).

#### V.B. Components of construal

Each construal pertains to a continuous period of time (T) and a spatial region (S). Within this spatio-temporal framework (see also Zwaan & Radvansky, 1998), there is a perspective (P). Sometimes, the perspective is that of a protagonist (Pr), but this is not necessarily the case. In each construal, there is a focal entity (F) and a relation (R), which is denoted by a verb or preposition. Often, there is also a backgrounded entity (B). The entities may have articulated features (f) that form part of the construal. These components are depicted in Figure 1.

INSERT FIGURE 1 ABOUT HERE

### V.B.1 Time interval

Construals involve events that take place during certain time intervals. It has been shown that comprehenders keep events active in working memory through extended time intervals (in the referential world), as long as the event is ongoing (Zwaan, Madden, & Whitten, 2000). Presumably, this is because comprehenders are engaged in continuous tracking of the referent situation (Kelter, Kaup, & Claus, submitted).

### V.B.2. Spatial region

Construals pertain to situations that take place in a spatial region. A region is defined here as a section of space delimited by the human senses and effectors. Often, these will be areas with boundaries that may occlude space outside the region from vision and limit the amount of sound and odor from the outside space, for instance rooms. Absent such obstructions, the range of the human senses provides the boundary of the region. Analysis of spatial layout perception provides an embodied tripartite classification of space (Cutting, 1997; Cutting & Vishton, 1995):

- personal space (around the observer and within arms' reach, with a 1.5 m radius);
- action space (an area around the observer in which he or she can walk quickly, talk, and throw things, with a 30 m radius);
- vista space (beyond 30 m from the observer).

A great number of studies have shown that the current location of a protagonist is more active in the comprehender's memory than other locations (e.g., Morrow, Bower, & Greenspan, 1987; Morrow, Greenspan, & Bower, 1989; Rinck & Bower, 1995).

### V.B.3. Perspective

Perspective is understood here as a point within a time interval and spatial region from which the referential situation is experienced. Thus, perspective is the spatio-temporal relation between the experiencer and the situation. Sometimes, a perspective is explicitly stated, e.g., “From the mountain top, the village looked tiny”, but often this is not the case. In such cases, content words may function as implicit perspective builders. For example, “The mouse approached the fence” implies not only a different distance between the focal and background entity than does “The tractor approached the fence,” it also implies a different distance between the experiencer and the situation (Morrow & Clark, 1988). Given that we cannot see mice from 200 yards, but we can see tractors and fences from that distance, the mouse-fence situation is presumably construed from a different vantage point than the tractor-fence situation. In this sense, one might argue that perspective on the referent is part of a word’s meaning.

Several components of perspective can be distinguished. The first one is location. A perspective is a location within a spatial region from which the situation is experienced. Sometimes, location is implied by a verb. For example, a verb like “come” as in “He came into the room” implies that the experiencer is situated at the endpoint of the action, i.e., in the room. In contrast, “go” in “He went into the room” implies that the experiencer is situated outside of the room. It has been demonstrated that comprehenders are sensitive to these verb-induced perspectival differences (Bower, Black, & Turner, 1979); in fact, this is true for young children (Rall & Harris, 2000) as well.

The second perspectival component is distance. This component was already addressed earlier in the example from Morrow and Clark, where experiencer-situation distance was implied by the relative sizes of the mouse and the tractor. In such cases, the human perspective is implicitly

assumed. For example, “squeak” implies a different distance between the experiencer and the sound source than “blast,” and the same is true in the visual domain for “mountain” and “molehill.” In other words, while grammatical, (5) sounds odd, and may therefore give rise to the inference that Eric is employing some listening device.

(5) Eric looked across the valley to see where the squeak came from.

Viewed this way, it may not be far-fetched to claim that perspective is part of the meaning of some words (see also Miller and Johnson-Laird, 1976, pp. 387-388, who note that the intrinsic region around an entity may be linked to perception, but is also related to social and functional constraints).

A third perspectival component is orientation. Again, man is the measure of all things. Our visual field is limited by the locations of our eyes at the front of our heads. Although eye movements and head rotations allow us to expand our visual field without moving our bodies, the comprehender’s default assumption is that something that can be seen is roughly ahead of us. Thus, the sentence about the tractor and the fence implies not only that the experiencer is a certain distance from the focal entity (the tractor) and the landmark (the fence), but also that this scene is roughly in front of him or her. Orientation is sometimes also determined vertically. For instance, “cloud” implies canonically that the experiencer is somewhere below it and “grass” implies that the experiencer is above it. Front/back is a more salient dimension for language comprehenders than is top/bottom, which is more salient than left/right (Franklin & Tversky, 1990). Borrowing terms from the anatomy literature, we can identify these dimensions as the coronal (front/back), transverse (top/bottom), and sagittal (left/right) plane. The transverse plane is more salient than the sagittal plane because our environment and our bodies are asymmetrical across the transverse plane (because of gravity), but not (so much) across the sagittal plane. It is clear that some entities have

stronger implications about vantage points than others. For instance, “cloud” presumably strongly implies that it is above the experiencer, whereas “rain” does this to a lesser extent. Most obviously, this is the case with nouns denoting parts of larger entities. For example, “branch” not only denotes a part of a tree (and as such can only be understood as the attentionally focused part of tree, see Langacker, 2001, for a similar line of reasoning), it also means that the part is (a) typically in the upper-half of the tree and thus (b) typically above the experiencer.

A different type of perspective is psychological perspective. The implied experiencer may have certain emotions, goals, and knowledge that the comprehender as an immersed experiencer may adopt. A psychological perspective should have an impact on which parts of the spatial region are selected as focal entities. It is beyond the scope of this contribution to further develop this topic here, but this is an important task for future research.

#### V.B.4. Entities and features

Grammatical markers (word order, case markers) signify what should be construed as the focal entity and what as the background. The focal entity is often the subject of the clause (Langacker, 2001), but this is not always the case. For example, in “We saw the castle on the hill” the object is the focal entity. Background entities are often signaled by prepositional phrases. Features are usually referred to by adjectives.

As noted earlier, words initially diffusely activate multiple overlapping functional webs. Empirical evidence on lexical-ambiguity resolution shows that initially multiple sense are briefly activated (Kintsch, 1988; Swinney, 1979). These findings are consistent with the IEF. However, the IEF makes finer-grained distinctions than amodal models. For example, a bird in the sky has a different shape than the same bird in its nest. Thus, in this case the shape of the focal entity depends

on its location, a fact that is not captured in amodal representations. Put differently, the shape is articulated during construal. The idea that a specific, but still schematic, representation of the referent is articulated during construal not only applies to referents activated by nouns, but also to those activated by verbs (Morrow & Clark, 1988; “approach” is represented differently in “The tractor approached the fence” and “The mouse approached the fence”), prepositions (“on” means something different in “The wallpaper is on the table” than in “The wallpaper is on the wall,” Sanford & Garrod, 1998 ), adjectives (“fast” refers to something different in “fast runner” than in “fast car” and to something different altogether in “fast typist,” Pustejovsky, 1995).

## V.C. Summary

To summarize, during construal, articulated, but schematic, experiential representations of referents are formed. It should be noted that these representations may be dynamic (Freyd, 1987), if the corresponding experiential trace is dynamic, e.g., that of seeing, hearing, and smelling a match being struck. Figure 1 summarizes the components of construal in an admittedly rather abstract format. The referential unit that a construal pertains to is an event, and event takes place at a certain time (T) and in a certain spatial region (S). Within this spatio-temporal framework (see also Zwaan & Radvansky, 1998), there is a perspective (P). Within this perspective are a focal entity (F), a relation (R), and a Background entity (B). Each of these three may have features (f).

## VI. Integration

### VI.A. Definition of integration

Once an event representation has been (partially) construed, the comprehender proceeds to the next construal. This is indicated in Figure 1 by the arrow leading from  $C_n$  to  $C_{n+1}$ . Relevant components of the previous construal(s) will provide part of the content of working memory, along with the functional webs activated by the current word(s) and will therefore influence the current construal. Integration refers to the transition from one construal to the next. The assumption is that these transitions are experientially based.

#### VI.B. Types of transitions

In discussing experiential transitions, it is helpful to use Langacker's (2001) notion of attentional frame because, as mentioned earlier, it construes language comprehension as the language-based modulation of attention to a described state of affairs. Extensive analyses are needed to catalogue the experiential transitions that might naturally occur in discourse. Awaiting such analyses, it might be useful to review a few frequently occurring types (see Millis, King, & Kim, 2000 for a first approach).

In descriptions of static scenes, the experiencer typically modulates attention and the associated transitions are perceptual, mostly visual, in nature. Typical transitions in scene descriptions are zooming, panning, scanning, and fixating, where each construal simulates the visual experience of an object, part of an object, or an object feature. Cognitive linguists have discussed these visual operations such as scanning at the sentence level (e.g., Langacker, 1990; Talmy, 1996). Consider the following passage from Sir Arthur Conan Doyle's novel The Hound of the Baskervilles (p. 57), in which Watson describes Holmes' and his arrival at the house of the Baskervilles. I have segmented the text into attentional frames "... , [A1] and the house lay before us. [A2] .... the centre was a heavy block of building [A3] from which a porch projected. [A4] The

whole front was draped in ivy [A5], with a patch clipped bare here and there [A6] where a window or a coat of arms broke through the dark veil.” A1 establishes the house as the focal unit. A2 zooms in on the central part of it. A3 zooms in further on a part of that central part, the porch. A4 zooms further in on the front of the porch, and A5 and A6 convey scanning of and fixations on parts of the focal entity of A4. Thus, the construals prompted by these attentional frames are visual experiences of the denoted entities and the transitions between them are common visual processes. Another type of attentional change is a switch from one sensory modality to another one. Recent findings suggest that such transitions incur processing costs, both in perception and in semantic processing (Pecher et al., 2003).

In descriptions of dynamic scenes or action sequences in which the experiencer is strictly an observer, the transitions are modulated by changes in the scene that attract attention. There is compelling evidence in the visual attention literature that changes in scenes are only detected if they are attended to and that changes more relevant to the scene are more likely to attract attention than peripheral changes (e.g., Rensink, O’Regan, & Clark, 1997; Simons & Levin, 1997). In both the description of static and dynamic scenes, it is often the case that attention shifts from the environment to an internal state of the experiencer, for instance an emotion (e.g., anger after seeing one’s office broken into), a cognitive state (e.g., confusion upon seeing an abstract work of art or a memory when smelling a certain smell), or a physical state (e.g., an urge to drink when hearing someone open a beer bottle). When the experiencer—whose perspective the comprehender as an immersed experiencer is invited to take—is an agent, actively changing the environment, an internal state (e.g., a goal) might transition into an action, which may transition into a change in the environment, which may transition into another internal state (e.g., frustration when the current state of the environment is not consistent with the goal motivating the action).

### VI.C. Factors influencing integration

Several classes of variables influence the ease with which a construal can be integrated with the current memory representation. First, there is concordance with human experience. Second, there is the amount of overlap between the evolving mental simulation (i.e., what Zwaan & Radvansky, 1998 called the integrated model) and the current construal. The more overlap there is among components, the easier integration should be. Third, there is predictability and fourth there are linguistic cues. The discussion will focus on the first two, as they are more unique to the IEF, whereas the last two are common across frameworks.

#### VI.C.1. Concordance with human experience.

It is perhaps not too much of a simplification to assume that the human experience involves continuity of time, space, and perspective. Confined to a single body and brain, we only experience the world from one vantage point at any moment in time. From this vantage point, we can perceive a region limited by the range of our sensory organs. If we want to change our vantage point, we need to move, thereby continuously updating our vantage point, using vestibular, proprioceptive, and optic flow information. If we want to learn more about an entity, we can visually scan it in a systematic sequence of fixations and saccades, we can move closer to it, touch, manipulate, and smell or even taste it.

The Continuity Assumption holds that comprehenders have the default expectation (not necessarily at a conscious level) that these continuities hold when comprehending discourse. However, one of the defining features of human language is displacement, the fact that we can communicate about events and things that are not in our immediate environment, or even about

things that are impossible (Hockett, 1959). Thus, language enables us to lift our communication out of our immediate environment in a way that gestures and animal communication do not. However, with this freedom come discrepancies between actual experience and experience conveyed through language. For example, in language we can abruptly shift from one time interval to a later or earlier one (e.g., by using a time adverbial such as “an hour later”). The same is true of location and vantage point. The continuity assumption proposes that this forces comprehenders to override their default expectations, which causes a momentary increase in processing activity.

Research has shown that violations of temporal continuity lead to an increase in reading times (e.g., Mandler, 1986; Zwaan, 1996) and increased electrical activity in left-prefrontal areas of the brain (Munte, Schiltz, & Kutas, 2000). Similarly, abrupt changes of location and vantage point may lead to temporary processing difficulty. The operative word here is “temporary.” Obviously, the utility of a communication system that leads to permanent disruptions of the comprehension process would be close to zero.

Comprehenders are adept at creating new time intervals, regions, and perspectives when instructed to do so. But why do discontinuities occur in the first place, if they lead to hiccups, however minor, of the comprehension process? This is because the goal of linguistic communication is often not to give a moment-by-moment account of some series of events. An appropriate way to phrase this idea in the context of the present framework is that language is used to modulate the comprehender’s attention on a referential situation in order to convey a point of view (not in the narrow sense of perspective) on that situation. The point of view dictates which entities, events, and features are relevant and which ones are not. Although the omission of point-of-view-irrelevant information leads to discontinuities in the referential situation, it is still

necessary. Aristotle (*Poetics*, trans. 1967) was clearly aware of this when he exhorted the dramatists of this time to use plot as the organizing structure and omit events irrelevant to it.

#### VI.C.2. Amount of overlap

The second major factor influencing ease of integration is the amount of overlap between the results of successive construals. The Overlap Assumption generally predicts that the fewer elements that change, the easier integration of the current construal should be with the contents of (long term) working memory. This prediction is of course not unique to the IEF. It appears in the early version of the Kintsch and van Dijk (1978) model in the form of argument overlap. Zwaan and Radvansky (1998) expanded on this notion and identified five dimensions of overlap at the level of the situation model: time, space, causation, motivation, and focal entity. The IEF adopts time, space, and focal entity, and adds perspective, background entity, and features. It deals differently with causation and motivation, as will be discussed later. Thus, according to the IEF, a construal can be more easily integrated with the evolving mental simulation when it has the same location, perspective, focal entity, and landmark as the previous simulation. The IEF also allows for the generation of more specific predictions. For example, it should be easier to add a feature to an entity than to add a new entity.

Taken to its extreme, the Overlap Assumption predicts that a construal, which overlaps completely with the previous one, should be the easiest to understand. However, as mentioned earlier, integration involves experientially based transitions from one construal to the next. With complete redundancy between consecutive construals, there is no obvious transition (although one could be inferred) and as such, comprehension should not be facilitated. Therefore, the optimum amount of overlap is somewhat less than complete overlap.

Consider Figure 1 again. In conjunction with the Concordance Hypothesis, the Overlap Assumption makes the general prediction that the relative processing cost incurred by a transition should in part be a function of the level of embeddedness of a construal components. Thus, feature changes should be less costly than entity or relation changes, which should be less costly than perspective changes, which should be less costly than time and region changes. However, ease of integration is also determined by predictability and the absence or presence of relevant cues. These factors are discussed next.

### VI.C.3. Predictability

Goals and causation have been demonstrated to play an important role in narrative comprehension (e.g., Graesser et al., 1994). Here, I simply assume that these are specific instances of the more general factor of predictability. As the history of psychology has shown time and again, if a sequence of events is experienced frequently, we tend to anticipate the second event when presented with the first. A cause-effect relation is an example of a frequently occurring sequence, an experiential trajectory. As such, it should be conceptualized as a dynamic representation in which a transformation of the focal entity occurs, for instance a change in shape, motion, direction, or color. Goal/plan structures (Schank & Abelson, 1977) are another example of experiential trajectories. These are typically longer than causal trajectories, but also facilitate integration. Subsequent construals are already anticipated (in part) before they occur.

We often experience effects before causes. We become aware of a disturbance in our environment and then look for or infer a cause. The cause is then either perceived in the case of a continuous force (e.g., the wind blowing a chair across the ice), or inferred in the case of a punctate force (e.g., the wind having snapped a branch). Also, according to the ideomotor theory of action

(e.g., James, 1890; Knuf, Aschersleben, & Prinz, 2001) a mental representation of an action's intended effect is assumed to be the cause of the effect. Thus, we are accustomed to construing effects before causes. In fact, in languages such as English, the effect is foregrounded syntactically, and not the cause (Talmy, 1976). In this light, the finding of Mandler (1986) that a sequence of events reported in reversed chronological order leads to comprehension difficulty unless the two events are causally related makes sense. Because in narratives cause-effect sequences are often familiar, syntactic cues are all that is needed for the comprehender to infer causality. However, in other cases, for example in science textbooks, a causal relation needs to be signaled explicitly in order to prevent the comprehender from interpreting the sequence simply as a temporal one. In such cases, causal connectives are used. They provide a cue to the comprehender to initiate a force-dynamic construal in the sense of Talmy (1988), rather than two successive "static" construals, as in the case of a temporal sequence.

#### VI.C.4. Linguistic cues

A fourth factor that influences integration is the use of linguistic cues. Syntactic cues such as word order or case cue the comprehender with respect to the elements of the construal. Word order or case provide cues as to what the focal and backgrounded entities are. Tense markers provide information on where the event should be placed on the timeline, whereas aspect markers indicate whether the event should be conceptualized as punctual, or temporally unbounded. The indefinite article indicates that a new entity is introduced, whereas the definite article is a cue to carry the entity over from one construal to the next. For example, the indefinite article in "A blue jay landed on the bird feeder" is a cue to activate the blue jay functional web and use the relevant aspects of it (visual and kinetic, presumably) in a simulation. In contrast, the definite article in "The

blue jay landed on the bird feeder" is a cue to maintain activation of the blue jay representation from a previous simulation (e.g., if the previous sentence read: "A blue jay flew down from the tree") as the focal entity (Givón, 1992). Prepositions can be viewed as instructions on the placement of and distribution of attention over entities in the mental simulation. For example in one reading of "on," this preposition places one entity above the other and makes it the focal entity, whereas "under" makes the lower entity the focal one. Cognitive linguists have documented a large number of lexical and syntactic cues for integration (e.g., Langacker, 1990; Givón, 1992; Goldberg, 1998).

## VII. Empirical evidence consistent with the IEF

There is a body of empirical findings that is consistent with the IEF—as well as with more general embodied theories of cognition (e.g., Barsalou, 1999; Glenberg, 1997)—but not with amodal models of comprehension.

### VII.A. Words activate of perception/action-related brain areas

Words activate brain areas that overlap with areas that are active when their referent is experienced. For instance, tool words have been shown to activate motor areas in the brain and certain animal words have been shown to activate visual areas (Büchel et al., 1998; Martin & Chao, 2000). Similarly, threat words, such as "destroy" and "mutilate" presented as part of a modified Stroop task, activated bilateral amygdalar regions to a greater extent than do neutral control words, thus implicating subcortical structures in semantic processing (Isenberg et al. 1999). The amygdala's role in emotional processing is well documented (e.g., LeDoux 1995). In addition, activation was found in sensory-evaluative and motor-planning areas, areas that are normally activated when the organism senses danger. This is all the more noteworthy given that the subjects

ostensive task was not comprehending words, but naming the color in which they were shown.

Results such as these suggest a strong link between words and experience.

#### VII.B. Action representations are activated during comprehension

The role of action in language comprehension has also been demonstrated in behavioral experiments. Glenberg and Kaschak (2002, experiment 1) presented subjects with sentences such as “He closed the drawer.” Subjects indicated whether or not the sentence was meaningful by pressing one of two buttons on a button box. There were three buttons, which were arranged away from the subject, rather than from left to right. At the beginning of a trial, the subjects held their hand on the middle button and moved it either away from or toward themselves to respond by pressing one of the other buttons. Responses toward the participant were facilitated when the sentence described an action requiring the hand to move toward the protagonist (e.g., opening a drawer) and responses away from the participant were facilitated when the sentence described an action away from the protagonist (e.g., closing a drawer). Moreover, in a subsequent experiment, Glenberg and Kaschak (2002) obtained a similar effect for more abstract sentences dealing with the transfer of information (e.g., “I told him the story”), providing evidence for the idea that basic perceptual and motor patterns may be metaphorically extended to more abstract situations via simple syntactic patterns (Goldberg, 1999). Spivey et al. (2000) recorded eye movements while subjects imagined or recalled objects that were not present in the visual display. In both cases, observers spontaneously looked at particular blank regions of space in a systematic fashion, in an apparent effort to manipulate and organize spatial relationships between mental and/or retinal images (see also Laeng & Theodorescu, 2002). Klatzky et al. (1989) observed priming between hand shapes and semantic judgments. For example, the sensibility of “Throwing a dart” was judged more quickly when

subjects had their hands in the appropriate shape for throwing darts than when not. Crucially, lexical associations were ruled out as a cause of this effect.

#### VII.C. Perceptual representations are routinely activated during comprehension

Recent evidence suggests that while a word is being processed, perceptual representations of its referent are being activated. Dahan and Tanenhaus (2002) showed that the activation of perceptual information is immediate. Participants listened to words while looking at pictures. Three pictures were presented at a time. One picture that matched the meaning of the word and two semantically unrelated distractors. Crucially, one of the distractors was similar in shape to the canonical shape of the target object, while the picture denoting the target object was in a different shape. For example, if the word was snake (or “slang” in Dutch), then the three pictures would be a picture of an uncoiled snake (target), a picture of a coiled rope (canonical shape distractor), and a picture of an unrelated item. Even before the participants had heard the end of the word, they were fixating more on the snake and the coiled rope. This suggests that shape information was accessed immediately. Further evidence suggests that an object’s shape is articulated by a sentence context. For example, participants respond more quickly to a picture of an eagle with its wings spread out after reading “The ranger saw an eagle in the sky” than after “The ranger saw an eagle in the tree” (Zwaan, Stanfield, & Yaxley, 2002). In this case, the location of the entity constrains its shape. As a consequence, any mental simulation of the situation would include the appropriate shape. Effects have also been found for the construal of object orientation. Thus, after reading “He pounded the nail into the wall,” participants responded faster to a picture of a horizontal nail than to one of a vertical nail, while the reverse was true after reading “He pounded the nail into the floor” (Stanfield & Zwaan, 2001).

Words denoting parts of objects activate visual representations of the larger object. Zwaan and Yaxley (in press a, b) presented pairs of words in the middle of a computer screen with one word appearing below the other. On critical trials, the word pairs were of the following kind:

(6a)                   ATTIC  
                           BASEMENT

or:

(6b)                   BASEMENT  
                           ATTIC

Participants made speeded judgments as to whether the two words were semantically related. The relative positions of the words either matched, as in (6a), or mismatched, as in (6b), with the relative positions of their referents in the larger object (a house in this case). Participants made faster relatedness judgments when there was a match than when there was a mismatch. This effect was not due to the order in which the words were read, because the effect disappeared when the pairs were presented horizontally. This result can be explained by assuming that the subjects engaged in a perceptual simulation to make the semantic judgments. In the context of our current argument, these results suggest that comprehenders construct integrated perceptual simulations based on just two content words, whereby the degree of overlap between the outcome of this simulation and the relative positions of linguistic symbols affected judgments. Thus, the presentation of two context words prompts a perceptual simulation in a language-processing task. Using a visual-field manipulation, Zwaan and Yaxley (in press b) showed that this mismatch effect is limited to the right hemisphere.

Expecting stimuli in the wrong sensory modality (e.g., expecting a visual stimulus when receiving an auditory one) incurs processing costs in perceptual tasks (Spence, Nicholls, & Driver,

2000). The same is true in semantic priming (Pecher, Zeelenberg, & Barsalou, 2003). Subjects evaluated properties of objects (e.g., LEMON-SOUR). Property verification times were slower when participants had to switch from one modality to the next (e.g., LEMON-SOUR followed by TOMATO-RED) than when they stayed within the same modality (e.g., STRAWBERRY-SWEET), even when lexical associations were controlled for.

#### VII.D. Comprehenders respond as if “in” the narrated situation

If the comprehender behaves like an immersed experiencer, the contents of working memory should reflect the accessibility of objects and events in the real world given our human sensory, attentional, and action-related limitations. Therefore, working memory should be more likely to contain representations of:

- present objects than absent objects;
- present features than absent features;
- close objects than distant objects;
- ongoing events rather than past events;
- current goals than past goals;
- visible entities rather than occluded entities.

Actions in the actual situation should also be constrained by what is read about the referent situation. Therefore,

- when not otherwise engaged in the actual situation, eye movements should reflect the vicarious experience of the referent situation (Laeng & Theoroescu, 2002; Spivey et al., 2000)

- when not otherwise engaged, manual movements should reflect the nature of the described situation (Glenberg & Kaschak, in press).
- when otherwise engaged, perceptual simulation of the referent situation might be interfered with (Fincher-Kiefer, 2001).

The assumption that is generally made in text comprehension research is that (a) information that is currently in working memory is more accessible (i.e., more highly activated) than information that is not and (b) when probed, more accessible information will yield faster responses than less accessible information. Therefore, if probe words associated with the contents of working memory are presented, responses should be facilitated. In accordance with this logic, various studies have demonstrated that the contents of working memory during comprehension reflect the nature of the described situation. Probe-word responses are faster when the probe refers to:

- a present entity rather than an absent entity (Anderson, Garrod, & Sanford, 1983; Carreiras et al. 1997).
- a present rather than an absent feature (Kaup & Zwaan, in press).
- a present object rather than a distant object (Glenberg, Meyer, & Lindem, 1987; Morrow, Bower, & Greenspan, 1989; Morrow, Greenspan, & Bower, 1987; Rinck & Bower, 2001).
- an ongoing event rather than a past event (Zwaan, 1996; Zwaan, Madden, & Whitten, 2000).
- a current goal rather than an accomplished one (Trabasso & Suh, 1993).
- a visible entity rather than an occluded one (Rapp & Horton, in press).

As mentioned earlier, there is also evidence that comprehenders assume the spatial perspective of a protagonist in the story (Bower, Black, & Turner 1979; Bryant, Tversky, & Franklin 1992; Franklin & Tversky 1990; Morrow & Clark 1988; Rall & Harris 2000).

### VIII. Accounting for propositional findings

In arguing the case for propositional representations, a primary goal has been to distinguish them from purely verbal representations (Kintsch, 1998, p. 69). The argument is that propositions have a different structure from sentences and thus that effects of propositional structure should be observable empirically. This has indeed been demonstrated. Take for instance an example discussed by Kintsch (1998): “The mausoleum that enshrined the tsar overlooked the square.” This sentence contains two propositions, [OVERLOOK, MAUSOLEUM, SQUARE] and [ENSHRINE, MAUSOLEUM, TSAR]. Participants who are given the cue “overlook” are more likely to recall “mausoleum” than “tsar” (Wanner, 1975). Similarly, there is more priming between words from the same atomic proposition (e.g., mausoleum, square) than between words from different propositions (e.g., square, tsar) (Ratcliff & McKoon, 1978).

These findings certainly suggest that the mental representation generated during comprehension is structurally different from the linguistic input. However, they do not necessarily provide evidence for amodal propositions. Specifically, it seems that the IEF would make similar predictions. The sentence presents two attentional frames giving rise to two construals. The first construal takes a perspective within the mausoleum and has the tsar as the focal entity. The second construal seems to imply a perspective outside the mausoleum and has the square as the focal entity. This suggests that there may be overlap between the information captured by amodal propositions and information captured by construals. In fact, there are theoretical arguments that predicate argument structures, such as propositions are grounded in perception and action (Hurford, in press) and that perceptual symbols can be thought of as modal propositions (Barsalou, 1999).

It is important to note that an experiential analysis is not redundant with an amodal propositional analysis. Specifically, an experiential analysis makes more subtle predictions than an amodal analysis. For example, the fact that the perspective changes from one construal to the next is not captured by an amodal analysis. For example, because of the Continuity Assumption, the IEF makes different predictions for “The mausoleum that enshrines the tsar overlooks the square,” where the perspective changes, than for “The mausoleum that enshrines the tsar has a marble floor,” where the perspective remains inside the building. For example, it should be easier to integrate the two construals in the latter case than in the former (Radvansky & Zacks, 1991). The amodal framework makes no such prediction.

Thus, it can be argued that the information captured by amodal propositions forms a subset of the information captured by a perceptual analysis. The many empirical studies discussed in previous sections show that the information not captured by amodal propositions cannot be safely ignored if we want to enhance our scientific understanding language comprehension. In this context, it is important to reiterate the point made by Kintsch and van Dijk (1978) that (amodal) propositions are “a convenient shorthand” for representing information. Indeed, as these authors and the many researchers inspired by them have shown, impressive results can be obtained by using this shorthand, but as is shown here, it would be a mistake to elevate the shorthand to the status of longhand.

### VIII. Accounting for abstract language

A common criticism of perceptual symbol theories is that, whereas it is relatively easy to see how they deal with concrete information, it is difficult to see how they deal with abstract information. In order to address this issue, Lakoff (1987) and others have advanced the notion of

metaphorical extension. For instance, as infants, we have learned to associate the positive physical feeling of warmth with physical proximity to our mother. By extension, we later learn to think about human relations in degrees of warmth. Similarly, it has been proposed that so-called “light” verbs, such as “put” and “take,” which are acquired early in life, are associated with syntactic constructions that carry a certain concrete causal meaning which underlies the meaning of other verbs used in the same construction (Goldberg, 1998). Talmy (1988) shows how his theory of force dynamics can be metaphorically extended for the analysis of arguments. Consistent with these ideas, experiential effects on the understanding of abstract language have been demonstrated in recent research (Boroditsky & Ramscar, 2002; Glenberg & Kaschak, 2002). However, it is clear that a great deal more needs to be done in order to arrive at a full experiential account of abstract language.

## X. Conclusion

The main thesis of this chapter can be summarized as follows: language comprehension is a vicarious experience brought about by three component processes: activation, construal, and integration. The IEF should be viewed as an engine that generates predictions about language comprehension. Like any good theory, it is falsifiable. For example, the general prediction that event sequences that are not consistent with experience are more difficult than event sequences that are can be falsified. Similarly, the prediction that the contents of working memory during comprehension should reflect the content of working memory during experience of the described situation can be falsified. These are rather general hypotheses. It should have become clear that the IEF is capable of generating far more specific hypotheses. Below is a non-exhaustive list of such predictions, segregated by the component processes of activation, construal, and integration.

### Activation

- Words or morphemes activate experiential representations that have much finer shadings than word senses and may include various shapes of the referent object , a perspective on the object.

### Construal

- Perspective is necessarily and therefore routinely encoded during comprehension
- Perceptible features of a referent object should be more activated than non-perceptible features.
- Transformations of focal entities that are inconsistent with human experience should be more difficult to understand than transformations that are not.
- Violations of object affordances will be more difficult to understand than non-violations (Glenberg & Kaschak, 2002).

### Integration

- Temporal, spatial, and perspective shifts inconsistent with human perception and action will be more difficult to understand than consistent shifts. There already is evidence for temporal and spatial shifts (see Zwaan & Radvansky, 1998 for a review), but not yet for perspective shifts.

By putting tests of predictions such as these on the research agenda, we should be able to arrive at a better understanding of language comprehension. Although the various components of the IEF require a great deal of further specification, I hope to have demonstrated that it is a useful first approximation to an embodied theory of comprehension.

XI. References

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. Quarterly Journal of Experimental Psychology 35A: 427-440.

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.

Boroditsky, L. (2000). Metaphoric Structuring: Understanding time through spatial metaphors. Cognition, 75, 1-28.

Boroditsky, L. & Ramscar, M. (2002). The Roles of Body and Mind in Abstract Thought. Psychological Science, 13, 185-188.

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

Bryant, D. J., Tversky, B. & Franklin, N. (1992) Internal and external spatial frameworks for representing described scenes. Journal of Memory and Language 31:74-98.

Büchel, C., Price C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. Nature, *392*, 274-277.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernandez, A. (1997) The role of verbal tense and verbal aspect in the foregrounding of information in reading. Memory & Cognition, *23*, 438-446.

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.

Cutting, J.E. (1997). How the eye measures reality and virtual reality. Behavior Research Methods, Instrumentation, and Computers, *29*, 29-36.

Cutting, J.E., & Vishton, P.M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (eds.) Handbook of perception and cognition, Vol 5; Perception of space and motion (pp. 69-117). San Diego, CA: Academic Press.

Dahan, D., & Tanenhaus, M.K. (2002). Activation of conceptual representations during spoken word recognition. Abstracts of the Psychonomic Society, 7, 14.

Damasio A. R. (1994). Descartes' error: Emotion, reason and the human brain. New York: Grosset/Putnam.

Damasio, A.R., (1999). The feeling of what happens: body and emotion in the making of consciousness. Harcourt Brace & Company.

Duchan, J.F., Bruder, G.A., & Hewitt, L.E., eds. (1995) Deixis in narrative: a cognitive Science Perspective. Mahwah, NJ: Erlbaum.

Ericsson, K. A. & Kintsch, W. (1995) Long-term working memory. Psychological Review, 102, 211-245.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. Journal of Experimental Psychology: General, 210, 339-357.

Fincher-Kiefer (2001). Perceptual components of situation models. Memory & Cognition, 29, 336-343.

- Franklin, N., & Tversky, B. (1990). Searching imagined environments. Journal of Experimental Psychology: General, 119, 63-76.
- Freyd, J.J. (1987) Dynamic mental representations. Psychological Review, 94, 427-438.
- Gernsbacher, M. A. (1990) Language Comprehension as Structure Building. Mahwah, NJ: Erlbaum.
- Gernsbacher, M. A., & Shroyer, S. (1989) The cataphoric use of the indefinite this in spoken narratives. Memory & Cognition 17, 536-540.
- Gerrig, R. J. (1993). Experiencing narrative worlds. New Haven: Yale UP.
- 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.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. Psychonomic Bulletin & Review, 9, 558-565.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987) Mental models contribute to foregrounding during text comprehension. Journal of Memory and Language 26:69-83.

Goldberg, A. (1998) Patterns of experience in patterns of language. In: The New Psychology of Language: Cognitive and Functional Approaches to Language Structure, ed. M.

Tomasello. Erlbaum.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997) Discourse comprehension. Annual Review of Psychology, 48, 163-189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994) Constructing inferences during narrative text comprehension. Psychological Review, 101, 371-395.

Half, H. M., Ortony, A., & Anderson, R. C. (1976). A context-sensitive representation of word meanings. Memory & Cognition, 4, 378-383.

Hockett, C.F. (1959). Animal 'languages' and human language. Human Biology 31, 32-39.

Horton, W.S., Rapp, D.N. (in press). Occlusion and the Accessibility of Information in Narrative Comprehension. Psychonomic Bulletin & Review.

Hurford, J.R. (in press). The neural basis of predicate-argument structure. Behavioral and Brain Sciences.

Isenberg, N., Silbersweig, D., Engelen, A., Emmerich, K., Malavade, K., Beati, B., Leon, A.C., & Stern, E. (1999). Linguistic threat activates the human amygdala. Proceedings of the National Academy of Sciences, 96, 10456-10459.

James, W. (1890). Principles of psychology. New York: Holt.

Johnson-Laird, P. N. (1983). Mental models: towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard UP.

Kaup, B., & Zwaan, R.A. (in press). Effects of negation and situational presence on the accessibility of text information. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Kintsch, W. (1998). Comprehension: a paradigm for cognition. Cambridge, UK: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Klatzky, R.L., Pellegrino, J.W., McCloskey, B.P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. Journal of Memory and Language, 28, 56-77.

Knuf, L., Aschersleben, G., & Prinz, W. (2001). An analysis of ideomotor action. Journal of Experimental Psychology: General, 130, 779-798.

Laeng, B., & Teodorescu, D.S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. Cognitive Science, 26, 207—231.

Lakoff, G. (1987) Women, fire, and dangerous things: what categories reveal about the mind. Chicago: University of Chicago Press.

Langacker, R.W. (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.

LeDoux, JE (1995). Emotion: Clues from the Brain. Annual Review of Psychology, 46, 209-235.

Levelt, W.J.M. (1989). Speaking: from intention to articulation. Cambridge, MA: MIT Press.

MacWhinney, B. (1999) The emergence of language from embodiment. In: MacWhinney, B. (ed.). The Emergence of Language. Mahwah, NJ: Erlbaum.

Mandler, J. M. (1986). On the comprehension of temporal order. Language and Cognitive Processes, 1, 309-320.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. Current Opinion in Neurobiology, 11, 194-201.

McRae, K., Sa, V. R., de, & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology: General, 126, 99-130.

Miceli, G., Fouch, E., Capasso, R., Shelton, J.R., Tomaiuolo, F., & Caramazza, A. (2001). The dissociation of color from form and function knowledge. Nature Neuroscience, 4, 662-667.

Miller, G. A., & Johnson-Laird, P. N. (1976). Language and Perception. Cambridge, MA: Harvard University Press.

Millis, K.K., King, A., & Kim, H.J. (2000). Updating situation models from descriptive texts: a test of the situational operator model. Discourse Processes, 30, 201-236.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. Journal of Memory and Language, 28, 292-312.

Morrow, D. G., & Clark, H. H. (1988). Interpreting words in spatial descriptions. Language and Cognitive Processes, 3, 275-291.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. Journal of Memory and Language 26,165-187.

Muente, T.F., Schiltz, K., & Kutas, M.(1998). When temporal terms belie conceptual order. Nature, 395, 71 - 73.

Paivio, A. (1971). Imagery and verbal processes. Holt, Rinehart, and Winston.

Palmer, S. E., Rosch, E., & Chase, P. (1981) Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), Attention and Performance, IX (pp. 135-151). Hillsdale, N.J.: Erlbaum.

Pecher, D., Zeelenberg, R.& Barsalou, L.W. (2003). Verifying the properties of object concepts across different modalities produces switching costs. Psychological Science.

Posner, M.I., & DiGirolamo, P. (1999). Flexible neural circuitry in word processing. Behavioral and Brain Sciences, 22,299-300.

Pulvermüller, F. (1999). Words in the brain's language. Behavioral and Brain Sciences, 22, 253-270.

Pulvermüller, F. (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. Progress in Neurobiology, 67, 85–111.

Pustejovsky, J. (1995). The generative lexicon. Cambridge, MA: MIT Press.

Radvansky, G. A. & Zacks, R. T. (1991). Mental models and fact retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 940-953.

Rall, J., & Harris, P.L. (2000). In Cinderella's slippers? Story comprehension from the protagonist's point of view. Developmental Psychology 36, 202-208.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997) To see or not to see: The need for attention to perceive change in scenes. Psychological Science, 8, 368-373.

Rinck, M., & Bower, G. H. (1995) Anaphora resolution and the focus of attention in situation models. Journal of Memory and Language, 34, 110-131.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. Memory & Cognition, 28, 1310-1320.

Sadoski, M., & Paivio, A. (2001). Imagery and text: A dual coding theory of reading and writing. Mahwah, NJ: Erlbaum.

Sanford, A. J., & Garrod, S. C. (1981). Understanding written language: Explorations in comprehension beyond the sentence. Chichester, UK: John Wiley & Sons.

Sanford, A.J. & Garrod, S.C. (1998). The role of scenario mapping in text comprehension, Discourse Processes, 26, 159 - 190.

Schank, R.C. & Abelson, R.P. (1977). Scripts, plans, goals and understanding: an inquiry into human knowledge structures. Hillsdale, NJ: Erlbaum.

Simons, D. J., & Levin, D. T. (1997). Change Blindness. Trends in Cognitive Sciences, 1, 261-267.

Spence, C., Nicholls, M.E.R., & Driver, J. (2000). The cost of expecting events in the wrong sensory modality. Perception & Psychophysics, 63, 330-336.

Spivey, M.J., Richardson, D.C., Tyler, M.J., Young, E.E. (2000). Eye movements during comprehension of spoken scene descriptions. Proceedings of the Twenty-second Annual Meeting of the Cognitive Science Society (pp. 487-492).

Stanfield, R.A. & Zwaan, R.A. (2001). The effect of implied orientation derived from verbal context on picture recognition. Psychological Science, 12, 153-156.

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. (1976). Semantic causative types. In: Shibatani, M. (ed.), The Grammar of Causative Constructions. Syntax and Semantics, Vol. 6, (pp. 43-116). New York: Academic Press.

Talmy, L. (1988) Force dynamics in language and cognition. Cognitive Science, 12, 49-100.

Talmy, L. (1996). Fictive Motion in Language and "Ception". In: P. Bloom, M. Peterson, L. Nadel, and M. Garrett (Eds.), Language and Space (pp. 211-276). Cambridge, MA MIT Press.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. Discourse Processes, 16, 3-34.

van Berkum, J.J.A., Hagoort, P.M. & Brown, C.M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. Journal of Cognitive Neuroscience, 11, 657-671.

van der Meer, E., Beyer, R., Heinze, B. Badel, I. (2002). Temporal order relations in language comprehension. Journal of Experimental Psychology: Learning, Memory, & Cognition, 28, 770-779.

van Dijk, T. A., & Kintsch, W. (1983) Strategies of Discourse Comprehension. New York: Academic Press.

Wanner, E. (1975). On remembering, forgetting, and understanding sentences. The Hague:

Mouton.

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

Zwaan, R.A., Madden, C.J., & Whitten, S.N. (2000). The presence of an event in the narrated situation affects its activation. Memory & Cognition, 28, 1022-1028.

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, a). Spatial iconicity affects semantic-relatedness judgments. Psychonomic Bulletin & Review.

Zwaan, R.A., & Yaxley, R.H. (in press, b). Hemispheric differences in semantic-relatedness judgments. Cognition.

Author note

This research was supported by grant MH-63972 from the National Institute of Mental Health. I thank Brian Ross, Katinka Dijkstra, Carol Madden, Rich Yaxley, Dave Therriault, and Mark Aveyard for helpful comments on previous versions. Author's address: Rolf A. Zwaan, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270, email:

[zwaan@psy.fsu.edu](mailto:zwaan@psy.fsu.edu).

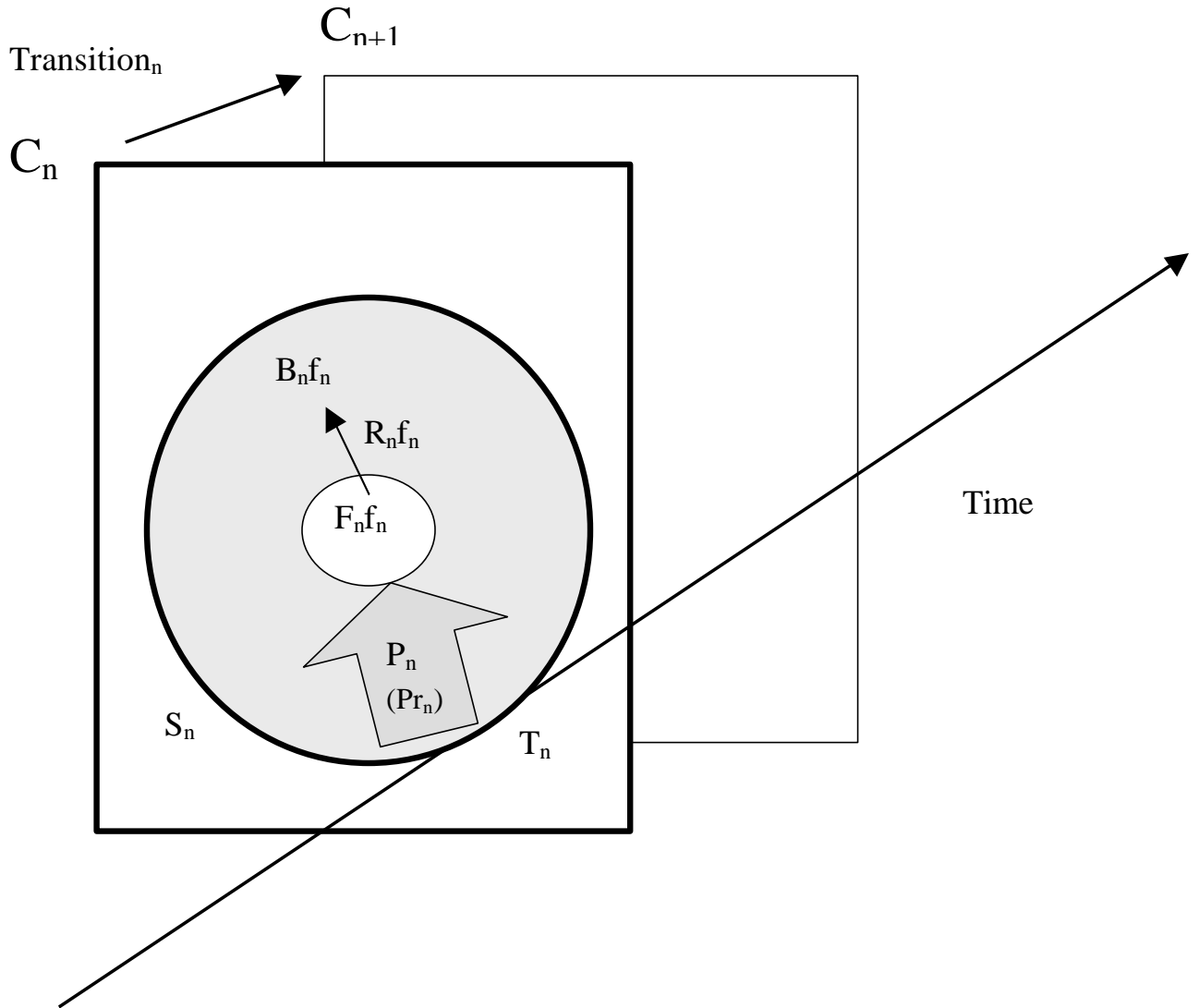
Table 1

Components of the comprehension process, the linguistic and representational units they operate on, and the referential units they pertain to.

Process	Linguistic Unit	Representational Unit	Referential Unit
Activation	Word/morpheme	Functional webs	Objects and actions
Construal	Clause/Intonation unit	Integrated webs	Events
Integration	Connected discourse	Sequence of Integrated webs	Event sequences

Figure caption

Figure 1: Structure and content of an attentional frame



C = construal

T = time

S = spatial region (personal, action, vista)

P = perspective

F = focal entity

R = relation

B = background entity

f = feature