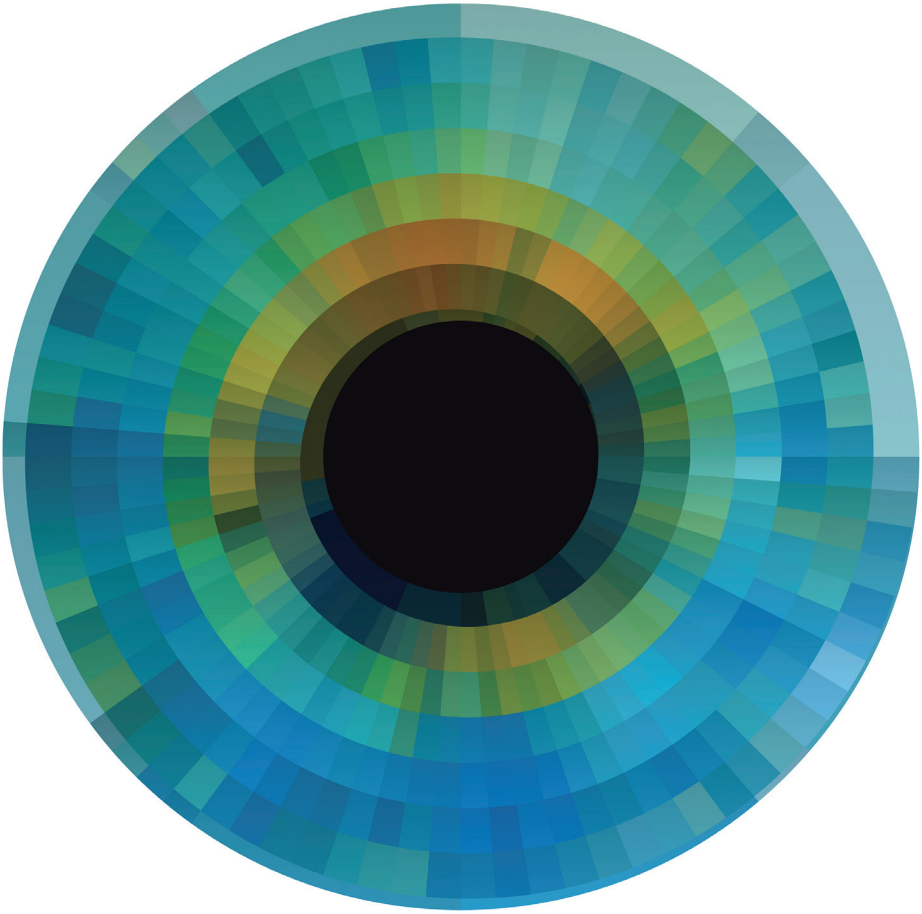


LANGUAGE IN THE MIND'S EYE

Visual Representations and Language Processing

Lisa Vandeberg

Language in the Mind's Eye:
Visual Representations and Language Processing

'Verbeelding' in taal: Visuele representaties en taalverwerking

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de
rector magnificus

Prof. dr. H.G. Schmidt

en volgens besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 16 november 2012 om 9.30 uur

door

Lisa Vandeberg
geboren te Venlo

ERASMUS UNIVERSITEIT ROTTERDAM

ISBN/EAN: 978-90-6464-594-5

Cover design: Wouter van der Sluijs
Lay out: Wouter van der Sluijs
Printing: GVO drukkers & vormgevers B.V.

Promotiecommissie

Promotor: Prof. dr. R.A. Zwaan

Overige leden: Prof. dr. T. van Gog
Dr. D. Pecher
Dr. K. Dijkstra

Contents

Chapter 1

Introduction and Overview

My favorite children's book was (and still is) Matilda, by Roald Dahl. The story is about Matilda Wormwood, an extraordinarily clever and sweet five year old girl who loves to learn and read. Unfortunately, her unpleasant parents are contemptuous of her inquisitiveness and talent, as is the headmistress of her school, Miss Trunchbull. While Matilda's parents force her to eat microwave dinners and watch loud game shows on TV, the child-hating Miss Trunchbull sows fear by locking children up in a device called the Chokey (a claustrophobic closet with spikes perforating the walls) or launching them across the schoolyard after swirling them around by their braids. Then, Matilda finds out she has psychokinetic powers and decides to use them to teach her parents and headmistress a lesson. The magic of this book was that it made me feel like being drawn away from reality and into Matilda's world: I could feel her eagerness to learn and her frustration with her parents, I could see her father's face and hear him shouting when she super-glued his hat to his head, and I envisioned what it looked like when she used her special powers to make crayons fly and write messages on a chalkboard to scare the life out of Miss Trunchbull. How is this possible? How can abstract symbols such as letters and words on a page come to life, engage you, create vivid images, and make you feel like you experience the described events yourself?

Mental representations

Mentally representing external or hypothetical worlds is at the basis of human thinking, learning, and functioning and has therefore received a great deal of interest in cognitive psychology, philosophy, and artificial intelligence. Comprehending a story, or any type of spoken or written language for that matter, requires the creation of such a mental representation, or a mental model. These mental models should go beyond representing the text itself and should rather represent what the text is about (Bower & Morrow, 1990; Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Thus, rather than representing the words and sentences that were written by Roald Dahl, readers should in some way represent the situation he described to truly understand and engage in the story. Ample empirical evidence suggests that people indeed create such *situation models*. Take, for example, the two sentence pairs that were used in a study by Bransford, Barclay, and Franks (1972):

Three turtles rested <u>on</u> a floating log, and a fish swam beneath <u>them</u>. (1a)
Three turtles rested <u>on</u> a floating log, and a fish swam beneath <u>it</u>. (1b)
Three turtles rested <u>beside</u> a floating log and a fish swam beneath <u>them</u>. (2a)
Three turtles rested <u>beside</u> a floating log and a fish swam beneath <u>it</u>. (2b)

Both sentence pairs have the same structure and contain the same words. However, the situation they describe is different. The described composition is identical for the first sentence pair (1a and 1b: fish beneath log, log beneath turtles), but different for the second pair (2a: fish beneath turtles, turtles beside log; 2b: fish beneath log, turtles beside log). Participants in this study were presented with a list of sentences that contained only one sentence of a pair (e.g., 1a). In a later part of the experiment they were presented with both members of a pair (e.g., 1a and 1b) and they had to decide which sentence they had memorized earlier. They did not have much difficulty discriminating between the members of pair 2 (which described different situations), but they experienced great difficulty remembering which member of pair 1 they had memorized (which described similar situations). Given that the linguistic information in both sentence pairs was identical, these findings demonstrate that memory for sentences is organized in terms of the described situation and not in terms of the described text itself.

Many studies since have shown that readers represent information about the described situation, for example, about the characters, objects, locations, actions, goals, intentions, time, and perspectives that are described in a text (e.g., Brunyé et al., 2009; Ditman, Holcomb, & Kuperberg, 2008; Egidi & Gerrig, 2006; Glenberg, Meyer, & Lindem, 1987; Magliano, Taylor, & Kim, 2005; Rapp & Taylor, 2004; Zwaan, 1999; Zwaan, Langston, & Graesser, 1995). However, knowing *what* we represent when processing language does not yet explain *how* we represent it. How did I represent the characters and objects in Matilda's story? In other words, how do people give actual meaning to the concepts they represent?

Meaning

For language to have meaning, words must in some way be connected to the things they refer to (Harnad, 1990). If they are not, they would merely be arbitrary formal symbols. Take the crayon from Matilda's example. The mere fact that this object is referred to as *crayon* by English speakers, *krijtje* by the Dutch, or *tiza* by the Spanish, shows that the language we use to refer to the real world is arbitrary. To understand the word *crayon* (or *tiza* or *krijtje* for that matter) we should in some way connect it to the actual object. To illustrate this point, consider the following thought experiment that was presented by Searle (1980). Imagine being in a closed room with a book, paper, and pencils. You receive Chinese characters through a slot in the door. You don't speak Chinese, but the instructions in the book explain how to process the symbols and how to produce different Chinese characters in response. You slide your response through the slot in the door, and receive new characters. This continues for a while. Suppose that you perform this task so convincingly that you convince a Chinese speaker at the other side of the door that you yourself are a native Chinese speaker. However, you did not understand anything of the 'conversation' you just had. Now consider that you have a dictionary that translates the Chinese symbols into words from your native language. This would surely increase your comprehension of the conversation. You would understand each symbol after you have translated it into your native language. However, the symbols in your native language are just as abstract and arbitrary as those in Chinese. How do you understand these symbols? What do they translate to that gives them meaning? Or specifically, how can a Dutch copy of Matilda make me experience another world whereas a Chinese copy would merely make me experience meaningless scribbles on a page? The proposed answer to this question is: By grounding these abstract symbols in experience.

In 1999, Barsalou published a highly influential paper called 'Perceptual symbol systems'. In this paper he proposed that there should be an analogue relationship between a symbol (e.g., the word *crayon*) and its referent (the actual object). He claimed that for knowledge to be analogous to the external world, the conceptual system that manipulates and stores this knowledge in the human brain should be perceptual in nature. To cite Barsalou,

> *'During perceptual experience, association areas in the brain capture bottom-up patterns of activation in sensory-motor areas. Later, in a top-down manner, association areas partially reactivate sensory-motor areas to implement perceptual symbols. The storage and reactivation of perceptual symbols operates at the level of perceptual components – not at the level of holistic perceptual experiences.' (p. 557)*

The first part of this quote means that the representations we create and retrieve when we think, learn, or process language are similar to the representations we created when we actually experienced the respective situation. Thus, giving meaning to the word *crayon* involves the reactivation of traces in memory that were laid down (grounded) during past experiences with crayons. Perceiving the world, interacting in it, and storing this information allows us to retrieve and represent what crayons look like, that they are used for writing, what it feels and sounds like when using them on a chalkboard, that they can break, that they are not edible, that they are often used by teachers, and so on. The second part of the quote suggests that parts of a representation can be flexibly recombined to create novel representations. This means that even if we have no physical experience with a certain situation (e.g., the flying crayons that were described by Roald Dahl), our experience with the individual components 'flying' and 'crayon' can be recombined which enables us to infer the approximate meaning and envision flying crayons.

Thus, for language to have meaning, mental representations must connect words to the things they refer to. This connection can take place if representations contain not only linguistic, but also perceptual (e.g., visual, auditory) and functional (e.g., motoric actions and affordances) information that is grounded in one's own experience of perceiving and interacting with the environment[1]. This idea was in itself not completely new. Theorists had already speculated about a role for visual and motoric information in the representation of concepts (Jackendoff, 1987; Lakoff, 1990; Langacker, 1986; Talmy, 1988). However, Barsalou (1999) specifically proposed that language cannot be informationally encapsulated in the human brain, but should rather interact with the neural systems that are involved in perception and action. In this case, the sensorimotor regions of the brain that are responsible for experiencing the environment should to some extent also be recruited when mentally representing (aspects of) this environment. With this view, he provided a theory about the mechanism that produces mental representations. This allowed researchers to start looking for neural and behavioral evidence to verify or falsify a role for these processes.

Evidence for visual representations

An abundant amount of evidence has now shown that people recruit perceptual information when they process language. This perceptual information appears to be stored and retrieved in a modality or property specific way (e.g., Barclay, Bransford, Franks, McCarrel, & Nitsch, 1974; Pecher, Zeelenberg & Barsalou, 2003). For example, Pecher, Zeelenberg, and Barsalou (2003) found that participants were faster to verify a certain property (e.g., auditory: 'A blender can be loud') when it was preceded by a property in the same modality (auditory: 'Leaves can be rustling') compared to when it was preceded by a property in a different modality (visual: 'A highway sign can be green'). This demonstrates that perceptual information from different modalities (vision, audition, taste, smell, touch, action) is represented differently. Let's now zoom into the visual modality, given that this might shed light on my experience of vividly 'envisioning' Matilda's adventures.

The first studies on the creation of visual representations during language processing demonstrated that readers were faster to identify pictures in which the visual features were congruent rather than incongruent with previously described events. For example, in an early study by Stanfield and Zwaan (2001), participants read sentences

1 *This experience may differ from person to person. For example, the representation you created when reading about Dahl's Matilda, Rowling's Harry Potter, Tolkien's Frodo, or Larsson's Lisbeth Salander may be quite different from the image the authors or the casting directors of their respective movies had in mind.*

like 'John put the pencil in the cup'. They identified subsequently presented pictures more quickly when these matched the implied orientation of the object (i.e., of a vertically oriented pencil) than when they did not (i.e., a pencil with a horizontal orientation). This finding suggests that readers activated the visual features of a pencil during reading and that these features affected subsequent responses to referent pictures. Since this finding, a variety of behavioral studies has demonstrated recruitment of visual information during language processing using different paradigms, stimuli, and dependent variables (e.g., Fincher-Kiefer, 2001; Horton & Rapp, 2003; Lupyan & Thompson-Schill, 2012; Richardson, Spivey, Barsalou, & McRae, 2003). Their results show that visual features of objects and actions such as shape, orientation, size, color, and motion direction are represented during language processing (e.g., Borghi & Riggio, 2009; Hirschfeld & Zwitserlood, 2010; Kaschak et al., 2005; Zwaan, Stanfield, & Yaxley, 2002) and suggest that this occurs routinely during reading and not as an artifact of the experimental paradigms that are used (e.g., Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009). Furthermore, fMRI evidence has shown that the brain areas responsible for visual perception are activated when processing language about visual situations (e.g., Desai, Binder, Conant, & Seidenberg, 2009; Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, 2007).

These findings provide evidence that the mental representations that people create are grounded in perception. Specifically, they show that situation models consist of visual information. This might explain how language can be so powerful that a comprehender actually feels like perceiving the described situation. In some sense, the findings suggest that readers actually do experience the wondrous world that is described in books. That is, they retrieve their own perceptions of the real world from memory upon reading the words, sentences, and paragraphs in the book. As the story unfolds, the different components of these experiences may together create a unique representation of the hypothetical world described in the book.

Note that visual information is not exhaustive of a conceptual representation. Mental representations include perceptual, lexical, semantic, and functional attributes and are shaped by people's world knowledge and beliefs (e.g., Murphy & Medin, 1985). Although visual information represents only one component of a concept, the interplay between visual information and language has proven to provide a very useful tool for studying human cognition. In the first part of this thesis we use this interplay to gain insights into how people organize and access multiple languages in their brain. In the second part we do not merely exploit visual information to study language, but we assess how language affects memory for visual information and how visual information affects memory for language.

Assessing the organization of multiple languages through visual referents

Knowledge of multiple languages in one brain can give rise to confusing, awkward, or funny situations. One might recall the famous Fawlty Towers scene in which John Cleese's character is determined not to mention the war to the German guests in his hotel. But of course, due to a head injury he does not succeed. After promising to bring the hors d'œuvres, he cannot resist to put up a German accent and inappropriately joke about 'orders ('*ôr dûrvs*') which must be obeyed at all times', hereby considerably upsetting his guests. This is one of many examples in which phonological overlap between two languages can create a funny play on words. However, form overlap between languages is not only a good source for word play. It can also be a useful tool to study how multiple languages are organized in one brain.

Many studies have used form overlap between languages to study how bilinguals access their lexical information. Most of these studies have used unbalanced bilinguals (those who have acquired their second language at a later age) and measured response times to words with and without form overlap across languages (e.g., Brysbaert, Van Dyck, & Van de Poel, 1999; Caramazza & Brones, 1979; Grainger & Beauvillain, 1987; Macnamara & Kushnir, 1971; Nas, 1983; Schulpen, Dijkstra, Schriefers, & Hasper, 2003; von Studnitz & Green, 2002). Abundant evidence suggests that bilinguals process words with cross-linguistic form overlap differently than words without this overlap. These differences are ascribed to influences from the other language. In other words, these results indicate that when bilinguals process words in a certain language, they might be unable to switch off the other language. This inability to suppress the other language is referred to as *language-nonselective access* or *cross-language activation*.

Although the differences in response latencies suggest an influence of the nontarget language in word processing, they cannot provide information on when and to what extent activation of nontarget language information takes place. *The visual world paradigm* (Cooper, 1974) offers a good solution, in that it provides an on-line measure of language processing. In this paradigm, the eye movements of people to real-world referents (objects on a screen) are recorded while they listen to sentences. Because language often refers to objects in the world and activates visual features of these objects, eye fixations to a relevant visual context can reveal the rapid mental processes involved in language processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

Several studies have used the visual world paradigm to assess cross-language activation in bilinguals. Because words in real-life almost never occur in isolation, examining bilingualism in a context is important for the generalization of the findings. Take the English word *note*, for example, which has form overlap with the Dutch word *noot*, meaning nut. Dutch-English bilinguals take more time to process *note* than to process a word that has no form overlap with Dutch (Dijkstra, Grainger, & Van Heuven, 1999), which suggest that they have difficulties deciding which *note* is implied (either the Dutch or English meaning). However, hearing 'I wrote you a note' might reduce this difficulty, because the context provides information about the word's language (English) and meaning (it must refer to something that can be written).

In **Chapter 2**, we describe the findings of studies that have examined cross-language activation in a sentence context. We then perform a study in which a sentence context not only provides information about the word's language and meaning but also about its word class. Consider the sentence 'Please note that the office closes at noon', for example. Here, the context implies an English verb rather than a Dutch noun. We use the visual world paradigm to examine whether this information might be sufficiently constraining for Dutch-English bilinguals not to be hindered by the Dutch meaning. If bilinguals would direct more visual attention to a depiction of the Dutch noun (e.g., a picture of a nut) than an irrelevant picture, this would indicate that cross-language activation even occurs across word classes in a sentence context.

In **Chapter 3**, we use cross-language activation as only one example of a broader principle termed *cognitive interactions*. Many researchers use visual referents in a visual world paradigm to assess whether the cognitive processes or representations that underlie viewing behavior interact, meaning that they affect each other. We argue that this definition needs to be further specified for eye movements to be able to actually reveal cognitive interactions. After reviewing the literature, we specify cognitive interactions as processes that should take place within an individual and that should be *temporally dependent* on one another. We argue that viewing behavior should be analyzed accordingly. To use a bilingual example: Evidence for cognitive interactions such as cross-language activation should reveal that the eyes of a bilingual are attracted to both the English and Dutch meaning of the word *note* in close relation to one another. We illustrate how other often used methods may falsely accept or reject interaction hypotheses and propose that assessing eye movement transitions, i.e., when the eyes move from one referent to another, can provide a solution to this problem. We demonstrate a method for assessing such transitions and explain how it can provide essential insights in human viewing behavior and its underlying cognitive processes.

Assessing the quality of visual representations through language

Up to now, we have placed people in a situated setting with relevant visual referents to study how they store and access linguistic information. In the second part of this thesis, we do not simply use visual information to study language, but we specifically ask how language can affect the quality of visual representations. We likely all have the subjective experience that language can affect the quality of the visual representations we make, just like my experience when reading Matilda. Reading a good book, listening to the lyrics of a song, overhearing an animated conversation, or retrieving memories with an old friend can bring vivid images to the mind's eye. But even less engaging language use can have this effect. I remember the time that the supervisor of this thesis recorded the stimuli for one of our experiments. He sat behind a microphone in the lab and read sentences off a piece of paper. He was aware that these Dutch sentences contained words with form overlap across languages and was instructed to read them in a neutral tone. At one point, he had to read out the sentence 'De jongen bedekte zijn haar met de pet' (meaning 'The boy covered his hair with the cap'). Right before pronouncing the word *pet*, his cross-language activation kicked in, which made realize that *pet* in English of course refers to a domesticated animal. This made him burst out in laughter, after which he explained that he envisioned a boy placing a small dog on his head. This of course made me envision the same scenario, after which we both were in a fit and needed 5 retakes to record the sentence properly. Luckily, not all sentences had this effect on us.

This example already illustrates that language at times evokes detailed and vivid visual representations, but not always. In **Chapter 4** we examine the vividness of visual representations by manipulating the transparency of pictures. People were presented with pictures of objects, after which they read scenarios in which these objects were mentioned. The presence and relevance of these objects in the described situation were manipulated. After reading the scenarios, the participants had to select out of two pictures the one that most resembled the picture they had seen before. However, one of the pictures was more transparent (less vivid) than the one they had seen and one was more opaque (more vivid). Now, imagine reading that Matilda really needs crayons to pull her prank on Miss Trunchbull, or reading a text in which the crayons are not relevant to her goal of scaring the headmistress. Furthermore, this text describes that she finds these crayons, or that there are no crayons. Research has shown that relevant or present objects are more readily available in one's situation model than irrelevant or absent objects (e.g., MacDonald & Just, 1989; Suh & Trabasso, 1993; Zwaan & Radvansky, 1998). However, notice that presence or absence of an object changes the visual appearance of a scene, whereas relevance or irrelevance of these objects does not. We hypothesized that representing the object in the situation model would increase the availability of visual information, and therefore increase the vividness of the mental representation in memory. We therefore expected that presence, but not relevance, would increase the visual vividness of the crayons.

Chapter 5 deals with the detail in which visual information is represented. Language naturally consists of different hierarchical levels of abstraction, for example, a superordinate level (fruit), a basic level (apple), and a subordinate level (granny smith) (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). We ask whether this hierarchy in linguistic specificity leaves its traces in the visual domain. Research shows that people encode general (category) information in terms of specific exemplars that are implied by the context (e.g., Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976). Consider the following two sentences:

The fruit forms an ingredient for cider. (1)
The fruit forms an ingredient for exotic salads. (2)

Both sentences mention the category fruit. However, the context of the first sentence biases towards the specific instantiation of an exemplar (apple), whereas the second sentence avoids this bias. When reading the word 'apple', people are better at retrieving sentence 1 than sentence 2 from memory, which suggests that they represented an apple when reading the first but not the second sentence. We use this principle to assess whether detailed line drawings (versus schematic line drawings) of, for example, an apple, provide better memory cues for sentence 1 than sentence 2. We hypothesized this to be the case, which would demonstrate that linguistic and visual specificity interact. We further examine whether such an interaction occurs during language processing (encoding) rather than after language processing (retrieval), which would suggest that representing exemplars is accompanied by greater visual specificity than representing categories.

Chapter 2

How verbs can activate things: Cross-language activation across word classes

Abstract

The present study explored whether language-nonselective access in bilinguals occurs across word classes in a sentence context. Dutch-English bilinguals were auditorily presented with English (L2) sentences while looking at a visual world. The sentences contained interlingual homophones from distinct lexical categories (e.g., the English verb *spoke,* which overlaps phonologically with the Dutch noun for ghost, *spook)*. Eye movement recordings showed that depictions of referents of the Dutch (L1) nouns attracted more visual attention than unrelated distractor pictures in sentences containing homophones. This finding shows that native language objects are activated during second language verb processing despite the structural information provided by the sentence context.

This chapter is published as: Vandeberg, L., Guadalupe, T., & Zwaan, R.A. (2011). How verbs can activate things: Cross-language activation across word classes. *Acta Psychologica*, 138, 68-73. doi:10.1016/j.actpsy.2011.05.007.

Introduction

When processing information in a particular language, does a bilingual automatically activate knowledge of both languages? Many studies from the domain of word recognition suggest this is the case. They have provided evidence for feature-based access to lexical information, rather than language-based access (e.g., Caramazza & Brones, 1979; de Groot, Delmaar, & Lupker, 2000; Grainger & Beauvillain, 1987; Schulpen, Dijkstra, Schriefers, & Hasper, 2003; von Studnitz & Green, 2002).

This feature-based approach, also referred to as language-nonselective lexical access, suggests that both languages are co-activated in a bilingual brain. In other words, this approach states that bilinguals cannot 'switch off' their other language. For example, when listening to Russian sentences containing a target word like *marku* (*stamp*), Russian-English bilinguals looked at a competitor picture of a *marker* (which has phonological overlap with the nontarget language English) more often than at competitor pictures unrelated to the target (Marian & Spivey, 2003; Spivey & Marian, 1999). Furthermore, an English word like the homophone *note* (with a similar sound but different meaning in Dutch, meaning nut) generates longer lexical decision times in Dutch-English bilinguals than a word that does not have phonological overlap with the other language (Dijkstra, Grainger, & Van Heuven, 1999). This is thought to be the result of competition for recognition between the English *note* and the Dutch *noot*. These results show that the native language can create interference in second language processing. This notion is supported by recent neuroimaging data which show language conflict at a neuronal level during second language processing, demonstrating that bilinguals are unable to suppress their nontarget native language to avoid interference (van Heuven, Schriefers, Dijkstra, & Hagoort, 2008).

In contrast to the studies mentioned above, in real life words are almost never processed in isolation; they usually appear in a context. Although it is necessary to understand language processing at a word level, this is not sufficient for a rigorous understanding of the bilingual language system. The context in which a word is processed is an important factor to take into account to provide ecological validity to our findings (Graesser, Millis, & Zwaan, 1997; Grosjean, 1998). A sentence context, for example, can provide information on the language membership, syntactic category (word class), or meaning of a word.

Not much research has been performed so far on how a sentence context modulates cross-language activation in bilingual lexical processing. Several studies that assessed reading times have demonstrated co-activation of nontarget languages in a semantic sentence context when target words were preceded by or embedded in *low-constraint* sentences, in which the sentence frame did not bias towards the target word (Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Elston-Güttler, Gunter, & Kotz, 2005; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009; Van Hell & De Groot, 2008). For example, Van Assche and colleagues (2009) found that a cognate word like the Dutch-English *oven* (with a similar word form and meaning across languages) generated shorter reading times than a non-cognate like *lade* (drawer) in a native language sentence context like 'Ben found an old oven/drawer among the rubbish in the attic'. Two studies that included *high-constraint* or predictive sentences (Schwartz & Kroll, 2006; Van Hell & De Groot, 2008) in which the sentence frame did bias towards the target word (e.g., 'The best cabin of the ship belongs to the *captain*'), failed to find facilitative cognate effects, suggesting that a context with a strong semantic fit can restrict nonselectivity and eliminate effects of the nontarget language on item processing. This notion is supported by a naming study (Li & Yip, 1998) in which highly predictive sentences generated faster naming responses to homophones than to control words, demonstrating that prior sentence context can affect the disambiguation of different homophone meanings in a bilingual context.

However, two recent studies have found evidence that a constraining context does not eliminate cross-linguistic effects, but merely reduces these effects. In a visual world study, Chambers and Cooke (2009) demonstrated cross-language lexical competition in both restrictive and nonrestrictive sentences, but found that fixations to a competitor picture (of a *pool*) dramatically reduced during and slightly after presentation of the noun (*poule*) in restrictive sentences (e.g., 'Marie va <u>nourrir</u> la *poule*'/'Marie will feed the chicken') as compared to nonrestrictive sentences (e.g., 'Marie va <u>décrire</u> la *poule*'/'Marie will describe the chicken'). In line with these findings, Libben and Titone (2009) found that strongly constraining contexts reduced the effects of interlingual competitors (and cognates) on reading times. Their findings suggest that between-language effects *can* occur, but are resolved rapidly. In their study, bilinguals read both high- and low-constraint sentences that contained cognates, homographs or matched controls. Eye movement recordings showed cognate facilitation and homograph inhibition in *early*-stage comprehension measures (e.g., the duration of the first fixation of the target word) of high-constraint sentence reading. No such co-activation effects were found in late-stage comprehension measures (e.g., the total duration of all fixations to the target word). These findings suggest that co-activation of languages can occur during the initial comprehension of constraining sentences, but is rapidly resolved at later stages of comprehension.

However, a sentence context does not only provide semantic information on the content of a word. Because a sentence context provides a structural frame in which words are interpreted, it also provides information on a word's syntactic category. For example, the English verb *bake* is phonologically identical to the Dutch noun *beek*, meaning creek. The placement of a word class ambiguous word like *bake* in a sentence context might also constrain interlingual lexical ambiguities. In other words, when processing a sentence in which the verb meaning of *bake* is used (e.g., 'the children bake cookies in the kitchen'), the inappropriate noun meaning (of creek) could become unavailable as a result of the word's placement at the verb position in the sentence. However, the few studies that have addressed how contextual constraints moderate lexical access in the bilingual language system, have all focused on co-activation of languages within a word class (nouns in particular). As a consequence, it is not clear whether the syntactic category information provided by a sentence context might affect bilingual lexical access across word classes.

The main objective of the present study was to explore whether language-nonselective access in bilinguals occurs across word classes despite the structural information provided by a sentence. Critical sentences contained word class ambiguous interlingual homophones at the verb location in English noun phrase–verb–noun phrase sentence structures. These homophones all constituted a verb in English (e.g., *step*) but a noun in Dutch (*step*, meaning scooter). It should be noted that a semantic ambiguity is inherently related to this syntactic ambiguity, in the sense that one needs to access the syntactic class of the word in order to access the meaning and vice versa.

To maximize the ecological validity of the experiment we presented the sentences a) auditorily, and b) in a situated visual world setting (Cooper, 1974). Because language often refers to objects in the world, eye fixations to relevant visual objects reflect the rapid mental processes involved in language, offering a measure of language processing as it unfolds over time (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The visual world we used consisted of two pictures depicting both nouns from the sentences (for practical reasons referred to as the agent and the patient, in respectively the first noun phrase and the second noun phrase), one picture depicting the noun counterpart of the homophone verb (also referred to as the between-language competitor), and an unrelated distractor picture. Hence, attention towards the competitor would have to compete with the sentences' actual meaning, which was depicted in both noun pictures of the agent and the patient. If nonselective access would occur across word classes despite the structural information provided by the sentence, this would be reflected by more visual attention towards between-language competitors relative to distractor pictures.

Method

Participants

Forty-one undergraduate students with Dutch as their native language (L1) and English as their second language (L2) took part in this study. Their proficiency in English was assessed by the Language Experience and Proficiency Questionnaire (LEAP-Q: Marian, Blumenfeld, & Kaushanskaya, 2007). The mean self-rated English skills were 7.4 (SD=0.87) on a scale from 1 to 10. The mean age at which these unbalanced bilinguals acquired English was 9.1 (SD= 2.8). All received course credit for their participation.

Stimuli

Sixty-six English noun phrase-verb-noun phrase sentences were constructed for this experiment. The 22 homophone sentences contained a monosyllabic interlingual homophone at the verb position, e.g., 'The boxers *step* into the ring' or 'The police officer *spoke* to the child'. The selected homophones were obtained from a pilot experiment[2]. The 22 matched control sentences were similar to the homophone sentences, the only difference being that the homophone verb was replaced by a 'monolingual' verb, e.g., 'The boxers *climb* into the ring' or 'The police officer *talked* to the child'. This was done in order to keep the semantic and syntactic sentential context the same. For an overview of the homophone and control verbs, see Appendix A. Furthermore, 22 filler sentences with an identical structure were created, e.g., 'The janitor walked to the store'. The visual world, consisting of four pictures in quadrants on the computer screen, was shown during each auditorily presented sentence. In the homophone condition the scene consisted of two noun pictures: The agent and the patient of the sentence (e.g., one of a boxer and one of a boxing ring), one between-language competitor picture (depicting the noun counterpart of the verb homophone, e.g., a scooter), and one unrelated distractor picture (e.g., of a traffic light), see Figure 1. The control sentences were accompanied by the same set of pictures. The positioning of the pictures from quadrants 1 to 4 was randomized across trials and across participants. The experimental sentences were counterbalanced between subjects so that each participant heard 11 homophone sentences, the 11 other control sentences, and all 22 filler sentences. The order of presentation was randomized within subjects. The sentences were recorded by a male native speaker of Dutch who had lived in the US for over 14 years, and sampled at 44.1 kHz. The sentences were rated by a native speaker of American English to ensure that they were pronounced correctly. The visual world was presented on the 21 inch display of the Tobii 2150 eye tracker, sampling at 50 Hz.

2 *In this pilot experiment, Dutch-English unbalanced bilinguals performed a Dutch auditory lexical decision task. Besides Dutch words and nonwords, the participants were presented with the English counterpart of a Dutch-English homophone. The 22 words that were mostly perceived as being Dutch (> 40% of the participants) were included in the experiment.*

Figure 1. An example of the Visual World Paradigm as used in the present experiment.

Procedure

The experiment consisted of three parts. First, participants performed the eye tracking experiment. They were seated with their eyes at approximately 60 cm from the display of the remote eye tracker. Participants were instructed in English. They received instructions for a so-called 'no-task task': They were not asked to perform any explicit task. Instead, the instructions were to listen carefully to the sentences while looking at the pictures displayed on the screen. This no-task task was chosen in order to avoid inducing unnatural processing strategies (see Altmann, 2004; Altmann & Kamide, 1999). Prior to each trial, a fixation point ('+') appeared on the screen for approximately 1 second. Next, the pictures were presented. After 3 seconds, the sentence was auditorily presented. The pictures remained on the display until 400 ms after the sentences' offset, after which the next trial started with a fixation point. After the last trial, participants were prompted to elaborate on whether they noticed anything particular about the relation between the sentences and the pictures.

The second part of the experiment consisted of a translation task. Forty-four English words were auditorily presented for the participants to translate into Dutch. Half of these words consisted of the homophones from the eye tracking experiment; the other half consisted of monosyllabic filler words. Again, the trials were randomized within subjects. This translation task made it possible to a) remove participants with a low proficiency in English and b) remove trials in which participants did not know the meaning of the English verb. The rationale behind this is that when English words are not known to a participant (either as a result of low proficiency of the participant or unfamiliarity with the word) these trials cannot properly reflect bilingual language processing; the participant would look at the competitor picture simply because he or she would not have the corresponding English word available. Therefore, the translation task was useful as a selection mechanism both at a participant and at an item level. Finally, the participants filled out the LEAP-Questionnaire.

Results

The data from three participants were removed due software malfunction. Furthermore, data from one item ('dote') and two participants with many incorrect responses to homophones in the translation task (homophone accuracy < .50) were removed from further analyses. Next, all trials with erroneous responses in the translation task were removed by item and by participant (12.31 % of the remaining homophone trials). Proportions of eye fixations to each picture were calculated for each participant in each condition (homophone versus control sentences) over successive time windows of 100 milliseconds, taking the verb onset as the reference point. The proportions of eye fixations across participants are presented in Figure 2.

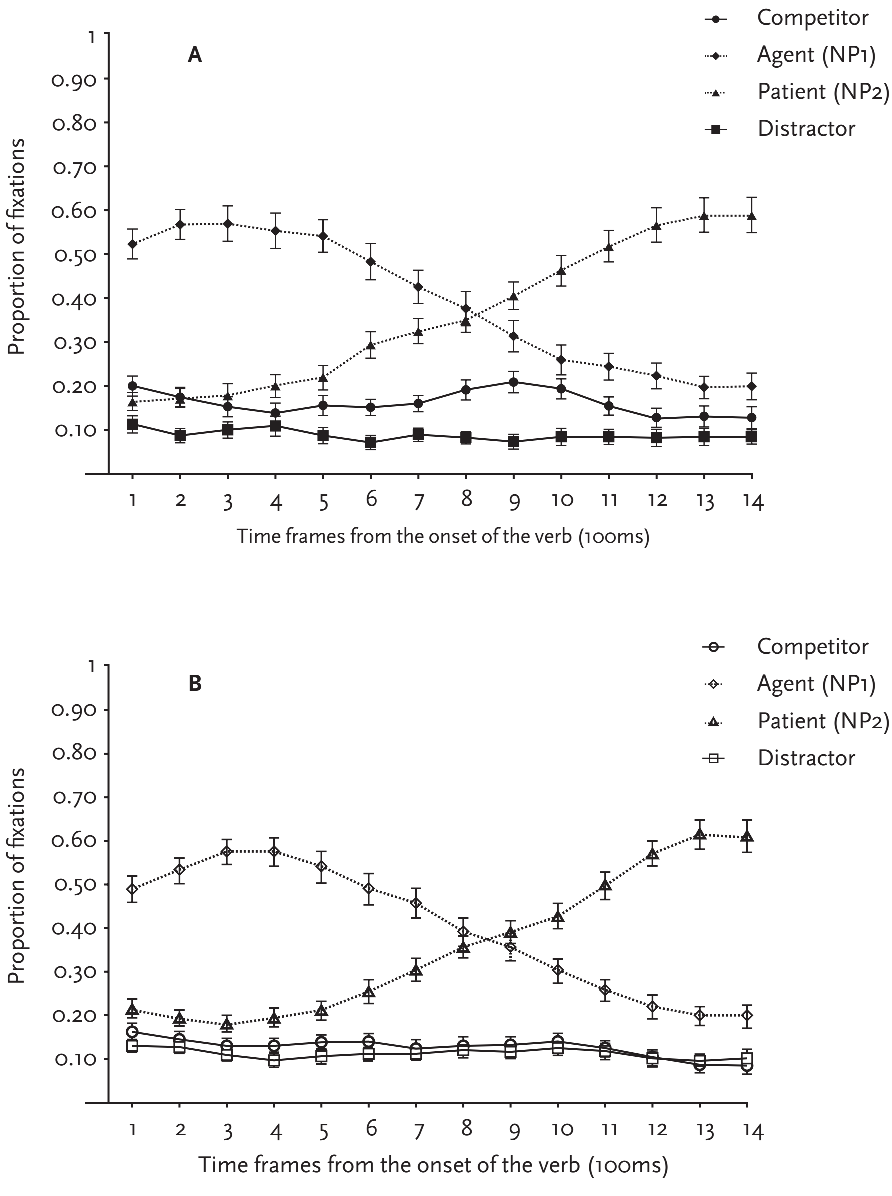


Figure 2. Mean proportions and standard errors of eye movements to each picture from the onset of the verb to 1400 ms after verb onset, for (A) the homophone condition and (B) the control condition. An example of the accompanying sentences is (A) 'The boxers step into the ring', (B) 'The boxers climb into the ring'.

A significant Picture (agent vs. patient)*Time interaction (F (13,58) = 13.92, $p < .001$, $\eta^2 = .76$) showed that early on in the sentence the agent received the highest proportion of fixations, whereas later on in the sentence the patient did; see the dashed lines in Figure 2A and B. This viewing pattern demonstrates that the participants were attending to the meaning of the sentence (or at the very least the meaning of the individual nouns), given that the agent always was a depiction of the first noun phrase and the patient a depiction of the second noun phrase.

The main purpose of this study was to explore whether language-nonselective access in bilinguals occurs across word classes in a sentence context. Note that both competitor and distractor pictures are irrelevant to the sentence context, but could in principle attract fixations. Therefore, the proportion of looks to competitors compared to the proportion of looks to distractors serves as an index of the degree of L2 activation. Repeated measures analysis revealed a significantly higher proportion of fixations to competitor pictures than to distractor pictures in the homophone condition (F (1,35) = 17.55, $p < .001$, $\eta^2 = .33$). These results show that, when processing a sentence like 'The boxers step into the ring', the picture of a scooter (/step/ in Dutch) received more visual attention than the distractor picture (depicting a traffic light). In other words, the bilingual participants activated the L1 noun counterparts of L2 homophone verbs during L2 sentence processing, reflecting co-activation of languages *across word classes*. Additional analyses showed that this effect could not have been the result of participants' possible awareness of the manipulation[3].

Given the finding that the homophone verb activated the native language in an L2 setting, we would only expect such an effect in homophone sentences as opposed to the control sentences. Indeed, similar analysis of the control condition showed no significant differences between competitor and distractor fixations (F (1,25) = 1.25, $p > .25$). When participants heard a sentence like 'The boxers climb into the ring', the

*3 In the post experiment session questioning the participants on the purpose of the experiment, seven participants indicated that they were aware of the presence of homophones in the stimuli. This subset of seven participants showed a significant difference in fixations to competitor and distractor pictures (F (1,6) = 14.64, p < .01, η^2=.71). These data show a strong effect, which possibly reflects a conscious search for a picture with phonological overlap across languages rather than first-pass lexical access to both languages, and which might have affected the overall analysis. However, after exclusion of this subset the significant difference in fixations to competitor and distractor pictures persisted (F (1,28) = 11.29, p < .005, η^2 = .29). Furthermore, when including the order of the trials as a factor (first two trials versus last two trials), this effect did not differ across trials (Picture * Trial interaction F (1,28) = 0.96, p > .3). These results indicate that the remaining participants indeed were not aware of the manipulation and suggest that the no-task task did not evoke a particular processing strategy of looking for a possible match between the pictures in these participants. Therefore, we conclude that the results reflect cross-language activation rather than a conscious search to the homophone counterpart.*

picture of a scooter did not receive more visual attention than the picture of the traffic light. This finding further indicates that the effect in the homophone condition cannot be the result of a difference in valence between the competitor and distractor pictures. Given that the identical visual scene evoked a different viewing behavior in different conditions, the difference in viewing patterns can only be attributed to the difference in linguistic input.

When looking into the homophone condition again, the difference in visual attention to competitor and distractor pictures changed over time (Picture (competitor vs. distractor)*Time interaction $F(13,23) = 2.18$, $p < .05$, $\eta^2 = .55$). To establish the time course of co-activation, a series of consecutive *t*-tests were performed to compare the proportion of fixations to competitors and distractor pictures per time window (see the uninterrupted lines in Figure 2a). Results of the two-tailed paired-samples *t*-tests at participant level showed that the proportions of looks to between-language competitors were significantly higher than the proportions of looks to distractor pictures (t_1 (35) > 2.12, $p < .05$) from 400 to 1100 ms after the onset of the homophone verb.[4] Item analyses showed a similar pattern in a slightly shorter time window (time frame 6, 8, and 9 t2 (20) > 2.32, $p < .05$, time frame 7 and 10 t2 (20) > 1.80, $p < .09$).

The increased visual attention to the between language competitor in homophone sentences must inherently have gone at the expense of attention to other pictures in the visual world. Analyses showed a significant difference in viewing behavior to both competitor and distractor pictures depending on the condition (homophone versus control sentences), as reflected by a significant Picture*Condition interaction (F (1,70) = 7.58, $p < .01$, $\eta^2 = .10$), see Figure 3. Furthermore, agent and patient fixations showed no such condition effects (Picture*Condition $F(1,70) = 0.03$, $p > .86$). These results suggest that between-language activation might have gone at the expense of attention to distractor pictures.

4 *Consecutive t-tests showed significant p values for the competitor-distractor comparison in time frame 1-2 and 5-11. An effect in proportions of looks was defined as the point at which five of more consecutive t-tests showed a significant difference (p <.05). This is a methodology frequently used in ERP research, e.g., Dehaene et al., 1998; van Schie, Mars, Coles, & Bekkering, 2004). The only time window in which (more than) five time frames were significant is that from 5-11, reflecting 400 to 1100 ms after the onset of the verb.*

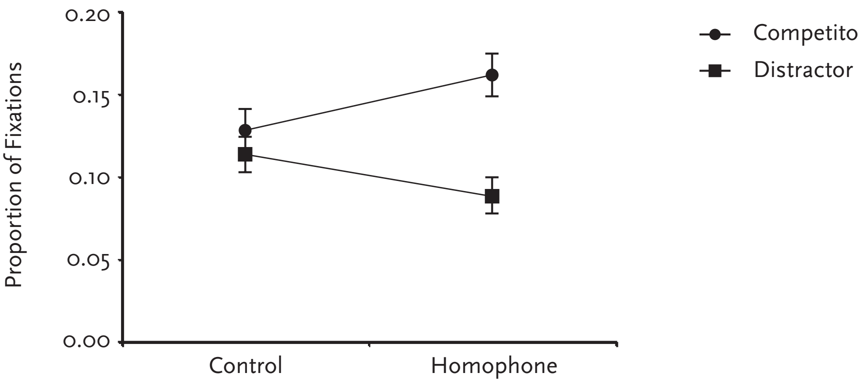


Figure 3. Mean proportions and standard errors of eye movements to competitor and distractor pictures for the control and the homophone condition.

Discussion

In a second language context, do bilinguals access knowledge of their native language across word classes despite being at verb position in a sentence structure? The results of the present study suggest they do. Second language sentences containing a homophone verb shifted visual attention towards the depictable noun counterpart in the native language, as was reflected in higher fixation proportions to competitor pictures than to unrelated distractor pictures. This activation across languages and word classes manifested *despite* a) the word being embedded in an L2 sentence frame that did not only provide information on the appropriate word's language membership and meaning, but also on its syntactic category, and b) competition of the homophone depiction with relevant agent and patient pictures. These findings provide additional evidence that bilinguals are unable to 'switch off' the native language when processing the second language.

The present study is, to our knowledge, the first to investigate whether word class information in a sentence context eliminates intrusion of the nontarget language during bilingual language processing. The only study explicitly investigating co-activation of languages across syntactic classes is that by Sunderman and Kroll (2006). In that study, English-Spanish bilinguals performed a translation recognition task on word pairs that either did or did not belong to the same syntactic category (e.g., *cara-card*, consisting of two nouns, versus *cara-care*, consisting of a noun and a verb). Results demonstrated cross-language lexical interference in same-class word pairs, but not in different-class word pairs. These results suggest that the influence of the other language was eliminated when crossing word classes. However, the present study does demonstrate cross-language effects across word classes. Both types of findings (selective versus nonselective lexical access across syntactic classes) do not necessarily contradict. It should first be noted that there are several differences between both studies, thereby compromising a direct comparison of the results. Compared to Sunderman and Kroll (2006), the present study used different stimulus contexts (sentences vs. word pairs), modalities (hearing vs. reading), and paradigms (eye movements vs. accuracy and response times). Importantly, also the timing of the measurements (during vs. after processing) differed. Therefore, it is possible that both studies were tapping different points in lexical access. It could be the case that the results by Sunderman and Kroll (2006) lacked the sensitivity to detect co-activation of languages due to measuring a single behavioral response *after* processing of the word pair. In contrast, the eye fixations in the present study provided more continuous measures *during* processing of the words in the sentence.

The notion of tapping into different levels of lexical access can also explain the seemingly contradicting results from the few cross-language studies on constraining contexts. As explained earlier, two studies on language-nonselectiveness in a sentence context suggested that a constraining context can restrict nonselectivity (Schwartz & Kroll, 2006; Van Hell & De Groot, 2008). On the other hand, two studies have demonstrated nonselective lexical access in constraining sentences, but only dramatically reduced (Chambers & Cooke, 2009) or during early stages of comprehension (Libben & Titone, 2009). The present study also demonstrated nonselective lexical access in a sentence context, and extended these findings across word classes. It should be noted that the type and extent of sentence constraints in the present study were different than those of the previously mentioned studies, given that the previous studies provided semantic cues on the appropriate meaning of the word in the preceding sentence context, whereas the present study mainly cued the syntactic category of the word by its placement in the sentence context. Again, both the findings of selective versus nonselective lexical access in a sentence context do not necessarily exclude one another. The studies supporting co-activation of languages (the present study; Chambers & Cooke, 2009; Libben & Titone, 2009) provided measurements during sentence processing, whereas the studies supporting language selectivity (Schwartz & Kroll, 2006; Van Hell & De Groot, 2008) provided a single measurement after sentence processing. Again, these findings indicate that the more fine-grained comprehension measures were sufficiently sensitive to reveal evidence of co-activation of languages.

Furthermore, the results from the present study suggest that only irrelevant visual information suffered from native language activation. Attention to between-language competitor pictures did not seem to occur at the expense of attention to the relevant noun phrase referents (agent and patient). In other words, the present data do show language nonselectivity but do not show impairment in processing the explicit meaning of the sentences. This finding can reconcile both viewpoints of evidence for co-activation during processing on the one hand and no evidence for co-activation at a decision level on the other hand: L1 activation might not exert sufficient influence to hinder L2 processing and, therefore, co-activation is not detected at a decision level.

Rather than reporting continued observations of language-nonselectivity, the status quo of bilingual research asks for specification of the factors under which nonselectivity is attenuated or reduced, in order to gain more insight in the bilingual language systems and improve models of bilingual language processing. The results of this single experiment are, of course, not conclusive about the role of syntactic information in language-nonselectivity and do not rule out that syntactic information can constrain initial consideration of the irrelevant language. Because the present study did not manipulate high versus low syntactic constraints, we cannot pinpoint whether the elevated visual attention for competitors demonstrate that initial lexical access is either not affected by word class information or whether it is attenuated despite modulation by word class information. The next step could be to present participants with low and high syntactic constraint sentences (in convergence with studies on low and high semantic constraints, e.g., Chambers & Cooke, 2009; Libben & Titone, 2009), in order to investigate this issue. Another step in specifying the hierarchy of the factors that affect language-nonselectivity could be to present both within-word class homophones and across-word class homophones within an experiment, or to manipulate both the syntactic and the semantic constraints of preceding sentences.

Despite the need for more studies, the present study has made a first step to investigate the role of word class information in a linguistic context on consideration of the nontarget language. The present results suggest that the constraining effect of syntactic category is late rather than early in cross-language processing. The influence of the native language materialized 400 ms after the onset of the ambiguous verb and lasted 700 ms. The planning of an eye movement usually takes approximately 200 (Hallet, 1986) to 300 (Matin, Shao, & Boff, 1993) ms after the onset of the referent's name. Given the processing constraints of the auditory referent in the present study, for example with respect to its meaning and language membership, a 100 ms delay compared to referents without these constraints is remarkably early. Furthermore, the 700 ms duration of the homophone effect is striking given competition with the pictures of the first and second relevant noun competitor.

Mapping of these results onto present word recognition models can only occur indirectly, because such models do not yet explicitly include information on the syntactic category of a word or specify exactly how linguistic and nonlinguistic context affect word recognition. However, models as the extended Bilingual Interactive Activation

(BIA+) model by Dijkstra and Van Heuven (2002) can provide a theoretical framework to carefully start interpreting the present findings. The BIA+ model specifies a bottom-up system for word identification that contains linguistic context information at a lower level, and a top-down system that regulates control and contains nonlinguistic context information at a higher level. It is assumed that the phonological overlap of the interlingual homophones in the present study induced competition between lexical candidates from both languages in the lower level word identification system. Next, word class information could become activated in the word identification system as a cue on the meaning of the different lexical candidates. In other words, analogously to the interpretation of Sunderman and Kroll (2006) the word identification system could initially be 'blind' to syntactic category. Sunderman and Kroll furthermore suggested that top-down demands from the task/decision system would use the syntactic information that is activated at a later stage to decide on the appropriate meaning of the lexical candidates.

As opposed to the study by Sunderman and Kroll (2006), the present study did not require explicit decision making, reducing the need to resolve lexical ambiguities top-down. However, it could be that the presence of a visual world with relevant and irrelevant referents might have invoked decision-making strategies, imposing a (somewhat less explicit) task demand as compared to explicit decision making requirements. This interpretation of the results in light of the BIA+ model is in line with the previous claim that L1 activation might not exert sufficient influence to hinder L2 processing at a decision level. However, the exact function of both systems in the BIA+ model with regard to syntactic category, linguistic context, and nonlinguistic context are still underspecified and in need of a broader empirical basis.

To conclude, this study demonstrated that sentences containing a lexically ambiguous verb (the interlingual homophone) resulted in visual attention to the inappropriate noun counterpart from the other language. Verbs in second language sentence processing activated irrelevant native language objects, demonstrating that language-nonselectivity persists across word classes in a sentence context.

Acknowledgements

We thank Bruno Bocanegra and Mark Adriaans for their contribution to this paper, and Marcel Boom and Christiaan Tieman from the Erasmus Behavioral Lab for their help with programming the experiment. Furthermore, we thank Judith Kroll, Wouter Duyck, Ana Schwartz, and an anonymous reviewer for their valuable comments on a previous version of the manuscript.

Appendix A

critical verb	control verb
ate	eats
bake	make
based	inspired
beat	hit
bet	wait
bill	charge
bow	kneel
build	find
court	seek
dose	prescribe
dote	cherish
float	hover
hack	cut
mail	send
mess	play
pet	stroke
snoop	investigate
snore	sleep
spin	rotate
spoke	talked
stain	pollute
step	climb

Chapter 3

Detecting cognitive interactions through eye movement transitions

Abstract

Many eye tracking studies are designed to reveal momentary (cursory) interactions between cognitive processes within participants. However, traditional analyses mostly use the viewing tendency of participants over trials (e.g., assessing average proportions of fixations to visual referents), rather than individual fixation patterns within trials (e.g., assessing consecutive fixations across visual referents). We argue that assessing temporal dependencies between eye movements to relevant referents is ideal for detecting cognitive interactions when participants are viewing a visual scene, whereas other often used methods may falsely accept or reject interaction hypotheses. We demonstrate how to analyze eye movement *transitions* with a multilevel markov modeling approach using a relevant experimental example (bilingual co-activation in a visual world paradigm), and discuss the practical applications and theoretical implications when analyzing transitions in any type of eye tracking data.

This chapter is under revision as: Vandeberg, L., Bouwmeester, S., Bocanegra, B.R., & Zwaan, R.A. (under revision). Detecting cognitive interactions through eye movement transitions.*

*The first two authors have equally contributed to this paper.

Introduction

To make sense of the world around us, we need to constantly process the sensory information that is provided by our environment. By studying how we process this continuous input, researchers attempt to understand the mechanisms that are involved in human cognition. For example, studying people's eye movements while they process spoken language, perceive visual scenes, and/or interact with their environment, can reveal the underlying processes that are called upon when performing these actions. For this purpose, eye tracking is a popular and widely used method in cognitive psychology. A wide variety of research domains have adopted the eye tracking methodology to study such topics as social interactions, cognitive development, language acquisition and comprehension, visuo-motor learning, memory, perception, visual cognition, and action. Several methods for assessing people's viewing behavior have been used, all of which study eye movements to regions of interest (pictures, words, or locations) in a visual scene.

The *visual world paradigm* (Cooper, 1974) is an exemplary method. This paradigm is based on the notion that looking to real-world objects while simultaneously listening to spoken language might reveal cognitive mechanisms involved in language processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Measuring the eye fixations to a relevant visual scene while sentences incrementally unfold provides a sensitive real-time index of how language is being processed. By recording eye movements to real-world referents (often pictures on a screen), researchers aim to study which representations are activated during language processing, when they are activated, and to what extent. Another well-known paradigm used with eye tracking is *visual search* (Treisman & Gelade, 1980; Williams, 1966, as described by Findlay, 2004). Here, participants are instructed to search for certain objects on the screen. By manipulating aspects of the visual scene (such as object features, or the number of distractors), researchers can, for example, investigate how bottom up stimulus features and top down volitional control influence visual search (e.g., Henderson & Ferreira, 2004; Theeuwes, 2010; Wolfe, 1994; also see Hartsuiker, Huettig, & Olivers, 2011).

Cognitive interactions

Researchers who study people's eye movements in visual world paradigms are often interested in whether activation of one system triggers activation or inhibition of another system, i.e., whether the cognitive systems that underlie viewing behavior *interact*. The idea of using eye movements to assess interactions on a cognitive level (referring to e.g., co-activation or co-inhibition of representations), is that it may result in competition between referents that is visible at the occulomotor level. A computational way of assessing such cognitive interactions is to assess the signature of latent cognitive processes on viewing behavior. For example, Stephen and Mirman (2010) have analyzed the distributions of gaze steps (i.e., the Euclidian distance between consecutive gaze positions) in visual search and visual world tasks and found that these distributions showed evidence for interacting processes that drive visual cognition (but also see Bogartz & Staub, 2011, for a substantial criticism). However, whereas most researchers aim to reveal interactions between certain predefined processes or mental representations, this computational approach is unable to specify the nature of the underlying cognitive systems.

This paper is concerned with a more cognitive approach that assesses how specific hypothesized systems interact. With this purpose, researchers usually define several regions of interest using visual referents to elicit fixations that are hypothesized to reveal specific underlying cognitive processes. Generally, they assess the fixation durations or fixation frequencies to each region of interest. Next, the average proportion of fixations is calculated within participants or within trials and they are compared for the visual referents of interest. We will argue that comparing the fixation proportions between referents may lead to substantial problems, because they can lead one to erroneously conclude the presence and absence of momentary interactions between cognitive processes.

Other than comparing fixation proportions, some researchers assess interactions by comparing fixation latencies to the referent of system A in the presence of a referent of system B, compared to a control condition in which referent B is not present in the display (see, e.g., the design explained by Barr, 2008). A delay of fixating referent A in the presence (versus absence) of referent B should indicate an interaction between systems A and B. An inherent disadvantage of this method is that these delays have to be measured across trials. That is, this method always compares a trial in which referent B was present to a different trial in which it was absent. As a result, assessing delays always occurs (a) across items, in which a participant responds to different items in different conditions, which does not allow the researcher to detect interactions within a certain participant on any certain item, or (b) within items but across trials, in which a participant responds to the same item in different conditions; in this case, one item is presented multiple times which is undesirable in, for example, language research. As we will argue throughout this paper, the assessment of momentary (cursory) cognitive interactions ideally occurs within participants *and within trials*.

Analyzing *transitions* between regions of interest circumvents these disadvantages of proportions and delays because they are measured within participants and trials. They can therefore provide the information needed to detect momentary interactions between cognitive processes, such as co-activation or co-inhibition of representations. Transition analyses (i.e., assessing dependencies between different states) have been performed in various eye tracking paradigms (e.g., Althoff & Cohen, 1999; Henderson, Falk, Minut, Dyer, & Mahadevan, 2000; Simola, Salojärvi, & Kojo, 2008) and are arguably preferable over fixation proportions (e.g., Tanenhaus, Frank, Jaeger, Masharov, & Salverda, 2008; Frank, Salverda, Jaeger, & Tanenhaus, 2009) for answering different research questions. In this paper, we will specifically demonstrate that transitions are ideally suited for assessing momentary cognitive interactions.

A relevant framework to demonstrate this is that of co-activated representations in a visual world paradigm. Because this framework leads to relatively simple predictions about first-order transitions, this example will be used throughout the paper. However, the points we make generalize to any type of eye tracking paradigm in which researchers assess interactions between systems that influence eye movements to visual referents within a single predefined event. First, we will outline the theoretical framework from our example and the issues that arise when conceptualizing cognitive interactions. Next, we will illustrate the importance of assessing individual transition patterns with two hypothetical situations.

Theoretical framework

Several studies that have adopted the visual world paradigm concluded that auditory linguistic input can lead to momentary interactions on different representational levels, for example in physical shapes (e.g., *snake – rope*, Dahan & Tanenhaus, 2005), semantic associates (e.g., *piano – trumpet*, Huettig & Altmann, 2005; Yee & Sedivy, 2006), lexical candidates within a language (e.g., *beaker – beetle*, Allopenna, Magnuson, & Tanenhaus, 1998; e.g., *candle – candy*, Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy, 1998, also see Huettig & McQueen, 2007), and across languages (e.g., *marku* - *marker*, Spivey & Marian, 1999). Most of these conclusions were based on fixation proportions that were averaged across participants, trials, and time (but also see Huettig & Altmann, 2005). For example, when fluent Russian-English bilinguals were instructed in Russian to manipulate a picture of a '*marku*' (meaning stamp in English), a higher proportion of fixations was directed at a competitor picture of a *marker* than at distractor pictures with no phonetic overlap in either language (Spivey & Marian, 1999).

Such differences in fixation proportions (see Marian, Blumenfeld, & Boukrina, 2007; Marian & Spivey, 2003; Marian et al., 2003 for similar findings) are taken as support for *co-activation of languages*, or cross-language lexical access. This is the leading view that when bilinguals are presented with a word that has (semantic, orthographic, or phonological) overlap across languages, lexical candidates from both languages are activated on a representational level which may result in competition on a decision level (e.g., Brysbaert, Van Dyck, & Van de Poel, 1999; Caramazza & Brones, 1979; Grainger & Beauvillain, 1987; Macnamara & Kushnir, 1971; Schulpen, Dijkstra, Schriefers, & Hasper, 2003; von Studnitz & Green, 2002, also see Marslen-Wilson, 1987 and McClelland & Elman, 1986 for models of speech perception). We use this theoretical framework to propose that, under the right conditions[5], a momentary interaction between systems (for example as a result of co-activation of languages) should result in *temporal dependencies* between critical fixations within a trial. We propose that transitions are ideally suited for detecting such temporal dependencies in support of cognitive interactions[6].

If a Dutch-English bilingual hears the English word '*mice*', the phonologically similar Dutch word form '*mais*' (meaning corn) may also be activated. Assuming that eye movements to relevant objects indeed reveal lexical activation, this co-activation of word forms should be reflected as more looks to both the English depiction (mice) and the Dutch depiction (corn) of the word, relative to the irrelevant distractors (pictures of e.g., leaves and apples). However, to demonstrate that both referents momentarily interact, not only should these target and competitor fixations occur, they should be *temporally contingent within a trial*. If the English and Dutch word forms are both activated upon hearing the word 'mice' (see Dijkstra & Van Heuven, 2002 for a model of bilingual language processing), and if the eye tracking device can detect this co-activation (see discussion section), the word 'mice' should evoke fixations to both the English and the Dutch depiction. Given that it is impossible to fixate two pictures simultaneously given the visual angle at which the pictures are presented, the way to tap into (continuous) temporal dependencies in lexical activation is to assess whether (discrete) critical fixations occur alternately for the critical referents.

Conceptual issues

Many eye tracking studies, such as those described in the theoretical framework, are analyzed by performing *t*-tests (ANOVAs) on fixation durations, fixation proportions, or fixation counts to each region of interest within a certain time interval, by modeling growth curves on fixation proportions (Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Mirman, Dixon, & Magnuson, 2008), or by performing binomial multilevel analyses on log odds (Barr, 2008; Frank, Salverda, Jaeger, & Tanenhaus, 2009). The latter analyses

5 *These conditions are discussed in section 'Future development and considerations' of the discussion.*

6 *Please note that it is not our intention to define co-activation or competition within a specific architectural framework. Though representations in a parallel system are by definition co-activated and representations in a discrete system are by definition not, these different architectures may be exhibiting exactly the same temporal dependencies in their overt viewing behavior. Nonetheless, it is very well possible to detect cognitive interactions in eye tracking paradigms when adopting a descriptive model. Researchers use visual world or visual search paradigms to assess whether fixating one picture is conditionally dependent on fixating another picture, i.e., to claim that a competitor attracts the eyes during target selection. In this paper we propose a measure for such conditional dependencies.*

are perfectly suited for establishing interactions as an aggregate property over events, e.g., within participants but across trials. However, many theories -including that of cross-lexical activation- specifically predict momentary interactions within events, e.g., within participants and within trials. Hypotheses based on momentary interactions are ideally addressed with transitions, because the description level of the analysis is in concordance with that of the theoretical claim. To illustrate this point, consider panels A and B from Table 1.

Note. This table illustrates the hypothetical fixation pattern of one participant in ten trials after onset of the target word in case of temporal contingency between critical fixations (panel B; transitions from region 1 to 2 and vice versa) or no temporal contingency (panel A; no transitions between regions 1 and 2 within trials). Numerals within a cell represent fixations on respective regions of interest. Each column represents a time interval; each row represents a trial (in the upper part of each panel) or averaged proportion of fixations across trials (in the lower part of each panel).

Table 1. Hypothetical viewing patterns.

A. No temporal contingency between critical regions 1 and 2.

		Timeframe (ms)				
		0-250	251-500	501-750	751-1000	Total
Trials	1.	3	2	2	4	
	2.	2	2	3	1	
	3.	1	3	2	3	
	4.	1	1	1	1	
	5.	3	1	1	1	
	6.	4	2	2	2	
	7.	2	2	4	1	
	8.	3	1	1	1	
	9.	4	4	1	1	
	10.	1	1	1	1	
Proportions	Region 1	0.3	0.4	0.5	0.7	0.48
	Region 2	0.2	0.4	0.3	0.1	0.25
	Region 3	0.3	0.1	0.1	0.1	0.15
	Region 4	0.2	0.1	0.1	0.1	0.13

B. Temporal contingency between critical regions 1 and 2.

		Timeframe (ms)				
		0-250	251-500	501-750	751-1000	Total
Trials	1.	2	1	4	1	
	2.	1	2	4	1	
	3.	3	1	1	2	
	4.	2	1	3	1	
	5.	4	2	1	1	
	6.	1	4	4	1	
	7.	3	1	2	4	
	8.	3	2	1	1	
	9.	1	3	1	3	
	10.	4	3	1	1	
Proportions	Region 1	0.3	0.4	0.5	0.7	0.48
	Region 2	0.2	0.3	0.1	0.1	0.18
	Region 3	0.3	0.2	0.1	0.1	0.18
	Region 4	0.2	0.1	0.3	0.1	0.18

Both panels reflect the hypothetical fixation pattern of a participant over time to four regions on a screen in ten different trials. Region 1 represents the referent for system 1 (e.g., the picture of the English *mice* from our example), region 2 refers to system 2 (e.g., the picture of corn, the Dutch *mais*), and regions 3 and 4 refer to distractor pictures that do not represent a relevant system (e.g., apples and leaves). In trial 1 from panel A, for example, the individual first fixated region 3, then made a transition to region 2, kept fixating this referent, and finally fixated region 4.

Panel A shows a situation in which analyzing the commonly used fixation proportions per time frame would lead to the conclusion that systems 1 and 2 interacted (or, in terms of our example, that hearing the English word form *mice* activated the Dutch word form *mais*, which corresponds to the picture of corn). This conclusion would be based on a higher proportion of fixations on critical regions 1 and 2 relative to distractor regions 3 and 4. However, the individual trials show that there were only three out of ten trials in which our hypothetical participant fixated both critical regions, and importantly, that these critical fixations were never sequential. In terms of our example this shows that there were no trials in which the participant made a transition between the English and Dutch referents upon hearing the target word (e.g., 'mice'). For this reason, these data do not provide convincing evidence that both systems (word meanings) were activated and their referents were in competition upon hearing the target word. However, analyzing fixation proportions by means of ANOVAS or growth curve models would in this situation falsely lead to the conclusion that both systems interacted.

Panel B, on the other hand, shows a situation in which the traditional analyses would result in the conclusion that there was no interaction between both systems, given the equal proportions of fixating the referent for system 2 (region 2) and the distractors (regions 3 and 4). However, seven out of ten trials show that the participant did not only fixate both critical pictures, but that these fixations were temporally contingent (switching from region 1 to 2 and 2 to 1). This would provide convincing evidence for the interaction between systems 1 and 2, or, in terms of our example, for cross-language activation. Essentially, if English and Dutch lexical candidates are co-activated upon hearing the word 'mice', this must be reflected by the temporal contingencies between target and competitor fixations. This example shows that, because fixation patterns vary considerably between individuals and over trials, analyses based on fixation proportions would falsely lead to the conclusion that there was no interaction between both systems.

For this reason, we argue that the test for a momentary interplay between two systems should measure fixation transitions between critical pictures relative to distractor pictures in individual trials. As illustrated by our hypothetical examples, it is not sufficient to conclude that the referents of both systems were fixated across trials or participants (i.e., over events). Rather, we should establish whether a fixation to the referent of one system is temporally dependent on a fixation to the referent of the other system (i.e., within events, e.g., during the trial 'mice'). Of course, as a result of visual search for the target, distractors will be fixated at times and switches from and to these pictures will occur. However, interaction between the two systems inherently implies temporal dependency, which should result in systematically more switches between regions 1 and 2 than regions 1 and 3, for example.

An ideal way to assess individual transition patterns is by means of *markov modeling*, because this technique assesses changes in fixations (states) over time, calculates transition probabilities, and properly analyzes the data by addressing the temporal dependencies of the switches (see Cristino, Mathot, Theeuwes, & Gilchrist, 2010; Hwang, Wang, & Pomplun, 2011; von der Malsburg & Vasishth, 2011, for different methods of assessing direct transitions or sequences in eye movements). Also, *multilevel* markov models are able to account for the multilevel structure of the data (trials are nested within participants). The effects that can be tested are either first-order transitions (assessing direct transitions from fixating one region of interest to the next, as is the case in this study) or higher-order transitions (reflecting a chain of two or more transitions to different regions of interest, i.e., the sequence of fixations). In the following section we will describe a study that is tailored to the present discussion, in order to illustrate the type of predictions that can be made in markov modeling and demonstrate how this technique can be used in analyzing eye tracking data.

Application

Method

Participants

Forty-eight undergraduate students at the Erasmus University Rotterdam participated in the experiment for course credit or payment. The description of the Experiment stated that native Dutch (L1) speakers with second language English (L2) could sign up for an English experiment testing the English proficiency in Dutch students. Their proficiency in English was assessed by the Language Experience and Proficiency Questionnaire (LEAP-Q: Marian, Blumenfeld, & Kaushanskaya, 2007). The mean self-rated English skills were 7.0 (SD = 1.15) on a scale from 1 to 10 (averaged over speaking skills, reading skills, and understanding spoken language). The mean age at which these unbalanced bilinguals acquired English was 9.4 (SD = 2.1). All had normal or corrected-to-normal vision.

Stimuli and apparatus

The auditory stimulus materials were derived from a pilot experiment that provided an empirical measure of phonological overlap. They were recorded by a male native speaker of Dutch who had lived in the US for 15 years, because a pilot experiment required the homophones to be pronounced both in English and Dutch. A native speaker of American English assessed the stimuli to ensure that they were pronounced correctly. We selected the data of 14 homophone-items from a pool of 43 English one- or two-syllable homophones (e.g., *mice*) and 43 English one- or two-syllable filler words (e.g., *grapes*). Each word was preceded by the instruction 'Click on the …'. The auditory stimuli were sampled at 44.1 kHz.

The visual world, consisting of four quadrants on the 21 inch display of a remote Tobii 2150 eye tracker, was shown during each auditorily presented sentence. The participants' eye positions were continually registered at a sampling rate of 50Hz. Each quadrant in the visual world contained one picture. In the homophone condition, the target picture (e.g., mice) was accompanied by a competitor reflecting the Dutch meaning of the homophone word (e.g., corn) and two distractor pictures, with no phonological overlap with the target (e.g., apples and leaves). Each picture occurred in one visual display only, and each visual display was presented once. The positioning of the pictures from quadrants 1 to 4 was randomized across trials and across participants. In addition, the order of trials was randomized across participants.

Procedure and design

The experimental session consisted of three parts. First, participants performed the visual world eye tracking task. They were seated in front of the eye tracker. After calibration, they read the English instructions on the computer screen. They were instructed to listen to the presented sentences carefully and then click on the picture as instructed in the sentence. The participants performed five practice trials that were similar to the experimental trials but included different words and pictures. If there were no further questions after the practice trials, the experiment started. Before each trial, a fixation cross and the mouse cursor were presented in the middle of the screen. When the eye tracker recorded 500 milliseconds of fixations to the cross, the trial started. The fixation cross disappeared and the visual world appeared on the screen simultaneously with the onset of the sentence. Each trial was terminated at the moment the participant clicked on a picture. 1000 ms after the trial had ended the fixation cross again appeared on the screen and the mouse cursor returned to its central position, leaving an equal distance between the mouse cursor and each of the four pictures at the onset of the trial.

Afterwards, participants filled out a Dutch translation of the validated Language Experience And Proficiency – Questionnaire (Marian, Blumenfeld, & Kaushanskaya,

2007) on the computer. This questionnaire was designed to assess the participants' language profile. At the end of the experimental session, participants performed a vocabulary test. During this test, they were auditorily presented with the English homophone target items to translate into Dutch. The trials in which the participants did not know the correct translations would be removed from the eye tracking data for further analyses, making the vocabulary test a useful selection mechanism both at participant and at item level. The entire experiment lasted approximately 45 minutes.

Exclusion criteria

Participants were excluded from the sample when they did not report English as being their real L2 (N=2), had extremely high error rates in the vocabulary test (above .50, N=9), or as a result of software malfunction (N=2). Furthermore, trials that were defined erroneously in the vocabulary test or had extremely long response latencies (over 3500 ms after target onset) were excluded from the analysis.

Predictions

The current hypothesis about cross-language activation (or any type of interaction between systems) should importantly be formulated in terms of transitions towards targets, competitors, and distractors within a person within a trial. For this purpose, markov modeling provides an ideal approach. The results of a markov model analysis can simply be summarized in a transition matrix containing estimations of the transition probabilities to each picture. Markov modeling allows us to test detailed predictions regarding the content of transition matrices if co-activation occurs versus if it does not occur. The transition matrix contains (1) the probability of staying fixated on a certain picture (referred to as 'staying probabilities', resembling the fixation proportions that are used in traditional analyses) and (2) the probability of switching from one picture to another (i.e., transitions between pictures, referred to as 'switching probabilities'). Given that we were interested in the pattern of probabilities in our model rather than their absolute sizes, we formulated all staying probabilities relative to the other staying probabilities (using capital letters), and all switching probabilities relative to the other switching probabilities (using lowercase letters).

Table 2 shows the hypothetical transition matrix, depicting the relative staying and switching probabilities in the case that languages are co-activated and the referents of both languages show temporal dependencies. The predictions are formulated as the probability of switching to a certain picture (at a certain time frame *t*, which is represented by the columns in the matrices) given a fixation on a certain other picture (at the preceding time frame *t-1*, as represented by the rows in the matrices). In other words, the hypotheses pertaining to co-activation are based on the previously attended picture and comparisons between switching probabilities are made on a row level in the matrices. The diagonal cells in the matrix depict the staying probabilities and the off-diagonal cells depict the switching probabilities. This prediction matrix can be mapped directly onto the actual transition matrix that results from the markov model analyses.

Table 2. Prediction matrix depicting the relative staying and switching probabilities for correct trials in the case of co-activation.

		Time frame = *t*			
		English	Dutch	Distractor 1	Distractor 2
Time frame	English	A	a	b	b
= *t-1*	Dutch	c	B	d	d
	Distractor 1	e	f	C	f
	Distractor 2	e	f	f	C

Note. Identical letters represent equal (expected) proportions. The diagonal cells (containing capital letters) represent the staying predictions; the off-diagonal cells (containing lowercase letters) represent the switching predictions. The critical hypothesized relationships are: a>b, c>d.

Previous studies (e.g., Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005; Spivey & Marian, 1999; Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy, 1998) have found that participants showed a higher fixation proportion on the target than on the other pictures, as a result of the task demands. Furthermore, they found that participants showed a higher proportion of fixations on the competitor than on the irrelevant distractor pictures, but that this fixation proportion was smaller than that on the target. This asymmetry is depicted in Table 2 on the diagonal axis, showing the highest probability of a participant fixating the English target and remaining fixated on the English target in the successive time frame (A), a lower probability of fixating the Dutch competitor and staying fixated there in the successive time frame (A>B) and the lowest probability of fixating one of the distractor pictures and staying fixated (A>B>C).

Importantly, English target and Dutch competitor fixations should be temporally dependent if both languages are co-activated, as reflected in the off-diagonal cells of the upper two rows. If the English target is fixated by the participant, the probability of switching to the Dutch competitor should be larger than the probability of switching to one of the distractor pictures (a>b). The probability of switching from the Dutch competitor to the English target should be larger than the probability of switching to a distractor (c>d). However, because the task instructed to click on the English target, we expect that every switch would be most likely made to the English target rather than any other picture. If a participant fixated a distractor, the probability of switching to the English target should also be greater than the probability of switching to the Dutch competitor or the other irrelevant distractor, because of the task demands (e>f). In terms of our critical hypothesis, this shows that the probability of switching from the Dutch competitor to the English target (relative to switching to a distractor, c>d), is not the most informative measure of a temporal dependency between the English and Dutch depiction. For this reason, the most critical comparison is that in row one of Table 2: Are people more likely to switch from the English target to the Dutch competitor than to a distractor (a>b)? This comparison is depicted in Figure 1.

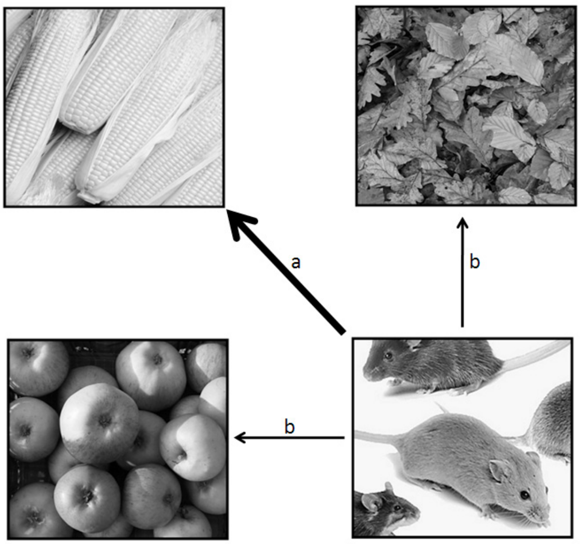


Figure 1. Exemplary trial of the Visual World Paradigm, in which the critical switching probabilities are depicted by the arrows.

Equations of the multilevel markov model

The multilevel markov model can be described by two multinomial regression equations which describe the distribution of the initial probabilities on the first time frame and the transition probabilities from time *t-1* to time *t*. The two sections below describe the two sets of multinomial regression equations that are involved in the current first-order multilevel markov models[7].

Initial probabilities

In the following equations, the letter *i* represents a participant (i=1,…, N), *k* represents a trial (k=1,…, K), *s* represents a fixation on a particular picture $[s=1,\ldots,4;\ \sum_{s=1}^{4} p(s_t)=1]$ and *t* represents a time frame of 100ms[8] (t=1,…T, in which T represents the last time frame of a trial, i.e., the moment of the click response which may differ across participants and trials). The log odds of the initial fixation probabilities ($t=1$) on a picture (s) by a participant (i) on a trial is defined as: $\pi_{is,t=1}$.

Because multiple trials are nested within participants, the fixation patterns of trials are dependent on participant. To correct for this dependency, coefficient $\pi_{is,t=1}$ is modeled as a function of fixed and random effects:

$$\pi_{is,t=1}=\alpha_{s,t=1}+\sum_{r=1}^{R}\alpha_{sr,t=1}\cdot r+g_{is,t=1}$$

$\alpha_{s,t=1}$ is the fixed intercept, reflecting the log odds of a fixation on picture *s* at $t=1$. $\alpha_{sr,t=1}$ reflects the fixed effect of a covariate *r* (r=1,…R, e.g., accuracy) on a fixation on picture *s* at $t=1$, and allows initial fixations to be different for different levels of the covariate (e.g., for correct versus incorrect trials). Finally, we have included a random intercept $g_{is,t=1}$ for individual participants, but no random slopes. This effect of participants is assumed to be normally distributed and centered around zero $\sim N(0,\sigma^2)$.

Because the coefficients $\pi_{is,t=1}$ sum to one over *s* pictures, we estimated *s*-1 (=3) parameters for each effect of $\alpha_{s,t=1}$, $\alpha_{sr,t=1}$, and $g_{is,t=1}$. The estimated log odds $\pi_{is,t=1}$ in the equation can simply be transformed into a transition probability $\phi_{is,t=1}$ which has a multinomial distribution:

$$\phi_{is,t=1}=\frac{\exp(\pi_{is,t=1})}{\sum_{s=1}^{S}\exp(\pi_{is,t=1})}$$

---

7 *For an introduction into multilevel modeling techniques see Baayen, Davidson, and Bates (2008). For categorical (multinomial) multilevel modeling in specific, see Agresti (2002) or Jaeger (2008), for example.*

8 *The data were aggregated to time frames of 100 ms, for reasons that will be thoroughly discussed in the discussion ('Future developments and considerations').*

Transition probabilities

In these equations, the letter u represents the fixation on a picture s at timeframe $t\text{-}1$ $[u=1,\ldots,4;\ \sum_{u=1}^{4} p(u)=1]$, whereas v represents the fixation on a picture s at timeframe t, $[v=1,\ldots,4;\ \sum_{v=1}^{4} p(v)=1]$. The log odds that a participant i on a trial k fixated picture v at time t, given that (s)he fixated picture u in the previous time frame $t\text{-}1$ is defined as $\pi_{i,v|u}$. This coefficient is in turn modeled as a function of fixed and random effects:

$$\pi_{i,v|u}=\gamma_{v|u0}+\sum_{l=1}^{L}\gamma_{v|ul}\cdot l+r_{i,v|u}$$

$\gamma_{0,v|u}$ is the fixed intercept, representing the log odds of a transition from picture u to picture v. $\gamma_{l,v|u}$ represents the effect of a covariate l (l=1,…L) on a transition from picture u to picture v. Intercept $r_{i,v|u}\sim N(0,\sigma^2)$ represents the random effect of individual participants on transition probabilities.

Because the coefficients $\gamma_{0,v|u}$, $\gamma_{l,v|u}$, and $r_{i,v|u}$ sum to one over v, we estimated s *(s-1) (=12) parameters for each effect of $\gamma_{0,v|u}$, $\gamma_{l,v|u}$, and $r_{i,v|u}$. The estimated log odds $\pi_{i,v|u}$ in the equation above can be transformed into a transition probability $\phi_{i,v|u}$ which has a multinomial distribution:

$$\phi_{i,v|u}=\frac{\exp(\pi_{i,v|u})}{\sum_{u=1}^{U}\exp(\pi_{i,v|u})}$$

The online technical Appendix (URL) contains an example of the data structure (Section A), the syntax (Section B), and the restrictions in the equations (Section C) we used for markov modeling in Latent Gold 4.5 (Vermunt & Magidson, 2005; 2008).

Results

The mean accuracy on the 14 homophone trials was 0.86 (SD = 0.14). The mean reaction time for the correct responses was 1913 ms (SD = 913) after target onset. Out of the false responses (corresponding to 55 trials), 92.73% was made to the Dutch competitor, whereas 1.82% was made to a distractor picture and 5.45% was not made to any picture at all (participants clicked next to a picture). Even though false responses should be removed from the data a priori, we did not remove them at this point in order to demonstrate how adding a covariate can affect the transition matrices. Markov analyses were performed on the data after 200 ms from the onset of the target up until the click response, because it takes approximately 200 to 300 ms to plan and fully execute an eye movement[9] (Hallet, 1986; Matin, Shao, & Boff, 1993). The analysis included the data of 35 participants on a total of 393 homophone trials, resulting in 5837 data points (measurements in time frames of 100 ms).

Fixation patterns were analyzed using the software package Latent Gold 4.5 (Vermunt & Magidson, 2005; 2008). Because the models were nested, we could use the difference between two models in log likelihood times 2 (2ΛLL) to test the difference in fit of the two models using a chi square distribution, in which the degrees of freedom are equal to the difference in parameters of the two nested models[10].

Hypothesis testing

In order to test the cross-language activation hypothesis we estimated a number of predetermined models (based on the theory) and compared their fit to the data. First, we performed a manipulation check to assess whether some important preconditions were met. Next, we built towards a model that performed the critical comparison for assessing co-activation between languages. Finally, we performed an additional test to validate our findings.

Manipulation check

The manipulation check was performed to test the assumption that both distractors were equally salient in our setup. Equal saliency would confirm that (a) the random positioning of the pictures was successful and (b) the contents of both distractor pictures were treated equal in this task. If this were not the case, any of the following results might have been caused by a failed manipulation rather than by cross-language activation[11].

The check consisted of the comparison of two models. In model 0, the distractor probabilities were based on freely estimated parameters and therefore allowed to vary. In model 1, the probabilities of the distractors in the transition matrix were constrained to be equal. If the distractors were equally salient in our setup we would expect that the fit of model 0 with the free distractor parameters was not significantly

---

9 *Altmann (2011) has reported the earliest meaningful saccade launch latencies (discriminating signal from noise) at about 100 ms. These launch times refer to the shortest possible time it takes to plan and initiate a saccade to the target picture, whereas we refer to the time it takes to plan and fully execute a saccade up until the eyes first land on the target.*

10 *Chi square tests are only appropriate for sufficiently large samples (see e.g., Agresti, 2002).*

11 *For example, temporal dependencies between target and competitor pictures might have been caused by unbalanced positioning of the pictures (e.g., target and competitor were always presented next to each other), whereas no temporal dependencies between target and competitor pictures might have been caused by a different saliency of the content of distractor pictures (e.g., one distractor picture was so salient that it artificially reduced competitor fixations).*

better than the fit of model 1 with the constrained distractor parameters. As expected, the fit of model 0 was not statistically better than the fit of the constrained model 1 ($2\Lambda LL$ (12) = 16.504, p = .17). This shows that the switching and staying probabilities did not differ for both distractor pictures, suggesting equal saliency for both distractors. For this reason, we took model 1 as the baseline multilevel markov model against which to compare the critical models.

Critical tests

Next, we estimated several models to perform the critical tests for co-activation. A discussion of these critical models is provided below, their fit to the data is summarized in Table 3.

Table 3. Fit statistics of the critical markov models.

Model	LL	#par	Comp.	2ΛLL	*p*
model 1: Baseline markov model	-4111.83	18			
model 2: Restriction dis = Du	-4147.93	8	1-2	72.21	<.0001
model 3: Restriction En->Du = En->dis	-4120.10	16	1-3	16.55	<.0001
model 4: Accuracy	-4071.67	34	1-4	80.31	<.0001
model 5: Accuracy, correct response: En->Du = En->dis	-4077.02	32	4-5	10.71	<.005

Note. En = English target; Du = Dutch competitor; dis = distractor pictures with equal probabilities. LL = log likelihood of the model; #par = number of estimated parameters; Comp. = comparison of the models; 2ΛLL = difference between the models in log likelihood times 2; p = the p-value.

Models 2 and 3: Evidence for co-activation in all trials?

In order to check the assumption that the viewing behavior with respect to the Dutch competitor differs from the viewing behavior to the distractor pictures (as should be the case when both languages are co-activated) we created model 2, in which the Dutch competitor probabilities were constrained to be equal to the distractor probabilities. If viewing behavior to the competitor differs from that to the distractors, the fit of model 2 should be significantly worse than the fit of model 1. The results indeed showed that the fit of model 2 was significantly worse than the fit of model 1 ($2\Lambda LL$ (10) =72.208, p <.0001). This demonstrates that the staying and switching probabilities of the Dutch competitor differed significantly from the distractor probabilities.

In model 3, the probability of switching from the English target to the Dutch competitor was constrained to be equal to the probability of switching from the English target to a distractor. The comparison of the fit of model 1 and model 3 is critical for co-activation. Because co-activation should be reflected in a temporal contingency between critical fixations – resulting in a higher probability of switching from the English target to the Dutch competitor than from the English target to a distractor, we expected model 3 to provide a significantly worse fit than model 1. Indeed, the fit of model 3 was significantly worse than the fit of model 1 (2ΛLL (2) =16.553, $p<.0001$). This demonstrates that the probability of switching from the English target to the Dutch competitor was significantly different from the probability of switching from the English target to one of the distractor pictures.

Models 4 and 5: Evidence for co-activation in correct trials

The results so far are based on all responses, including the incorrect click responses. A priori, incorrect trials might obscure our findings because the majority of the incorrect click responses (92.73%) was made to the Dutch competitor. Importantly, inaccurate click responses to the Dutch competitor will coincide with a higher transition probability from any picture to the Dutch depiction. Because transitions from the English to the Dutch depiction provide the critical test, incorrect trials might compromise this test and lead to a false positive. Therefore, it is necessary to assess whether a pattern of co-activation occurred *on the correct trials*. This way, a demonstrated temporal contingency between fixations on the English and Dutch depictions could not be the result of more switches to the Dutch depiction when making an incorrect response.

Model 4 builds further on our baseline model 1, which provides the best fit to the data so far. Model 4 included a fixed effect for the independent variable *accuracy*, which indicates whether people gave a correct (value 1) or incorrect (value 0) response on a trial. Because it is highly plausible that incorrect click responses coincide with very different fixation patterns than correct click responses, we expected model 4 to provide a better fit than model 1. We compared the fit of model 4 with the fit of model 1 and concluded that adding the covariate accuracy lead to a significant improvement in model fit, 2ΛLL (16) =80.311, $p < .001$. This demonstrates that correct responses to the English target resulted in different transition patterns than incorrect responses.

Next, in model 5 we constrained the probability of switching from the English target to the Dutch competitor to be equal to the probability of switching from the English target to a distractor when participants made a correct response[12]. Now, the

12 *Because we did not hypothesize the incorrect trials to provide viewing patterns that were relevant with respect to co-activation, we did not restrict probabilities in these trials.*

crucial test for cross-language activation is the comparison of the fit of model 4 and model 5. Importantly, the fit of model 5 was significantly worse than the fit of model 4 ($2\Lambda LL\ (2) = 10.706$, $p < .005$), which demonstrates that the probability of switching from the critical English target to the Dutch competitor was different from the probability of switching from the English target to a distractor when the participants made a correct response.

Table 4 shows the transition matrix belonging to model 4. The probabilities for the correct responses are presented in the upper panel, the probabilities for the incorrect responses are presented in the lower panel.

Table 4. Estimated transition matrix of the staying probabilities and switching probabilities.

			Time frame= *t*			
			English	Dutch	Dis1	Dis2
	Correct	English	.908	.040	.026	.026
		Dutch	.146	.749	.053	.053
		Dis1	.174	.083	.664	.080
Time frame= *t-1*		Dis2	.174	.083	.080	.664
	Incorrect	English	.865	.080	.027	.027
		Dutch	.026	.916	.029	.029
		Dis1	.077	.134	.724	.066
		Dis2	.077	.134	.066	.724

Note. The probabilities in each row add up to 1. The probabilities of the eye gaze staying fixated upon a certain picture are always greater than the probabilities of switching between two pictures. This is a result of calculating the transition probabilities per time frame of 100 ms (also see discussion section). It is impossible to switch from one picture to another every 100 ms, given that it takes 200-300 ms to plan an eye movement. Therefore, the staying probabilities can only be interpreted relative to the other staying probabilities, and switching probabilities relative to the other switching probabilities within a row, as pointed out previously in the prediction matrix.

Comparison of the transition matrix from the correct trials in Table 4 to the prediction matrix in Table 2 reveals an identical match. The observed transition matrix shows that when participants fixated the English target, they were more likely to switch to the Dutch competitor (.04) than to one of the two distractor pictures (.026). This is supported by the fact that model 4 provided a significantly better fit to the data than a model in which the probability of switching from English to Dutch was constrained to be equal to the probability of switching from English to a distractor (model 5; $2\Lambda LL$ (2) =10.706, $p < .005$). This finding demonstrates that fixations to the critical English target and Dutch distractor were temporally dependent in correct trials. This convincingly supports the notion that word forms from both languages were co-activated within trials in which participants responded correctly to the English target picture.

Though the difference between the Dutch switching probability of .04 and the distractor switching probability of .026 is significant, it may seem small in overall magnitude. However, it shows that out of all transitions that are made from the English picture (1-.908=.092), an estimated 43.5% (.04 out of .092) went to the Dutch competitor whereas only 28.3% (.026 out of .092) went to one of both distractor pictures. Thus, the relative differences between the transition probabilities are substantial (43.5% vs. 28.3%), even though the absolute values of the probabilities may suggest otherwise.

We did not have a priori hypotheses about temporal dependencies on the incorrect trials. However, as the transition matrix in Table 2 illustrates, there seem to be no critical temporal contingencies for the incorrect trials. For the incorrect responses to the Dutch picture, the critical comparison would lie in the probability of switching from the Dutch competitor to the English target compared to Dutch-distractor switches (analogous to the reasoning for the critical comparison in correct trials). Given that the probability of switching from the Dutch to the English depiction (.026) is numerically even lower than the probability of switching from the Dutch competitor to a distractor (.029), there is no evidence for cross-lexical activation on the incorrect trials.

Validation of the models

The latter models (4 and 5) have included a predictor on a trial level, namely accuracy. It is possible to include predictors on all levels of the multilevel markov model. For example, it is highly plausible that fixation patterns change over time during a trial. Fixations at the end of a trial should likely be directed at the picture that is selected for the response, whereas fixations before response selection and execution should likely be more diverse. To examine this in detail, it is possible to add a predictor on a time level within a trial. For the sake of validation, we have included this predictor in an additional analysis.

Model 6 included the independent variable *time interval*. This model was based on model 1 and not on model 4 that included accuracy, because we did not have a priori expectations of an accuracy*time interaction and this additional model merely served a validation purpose. The variable was divided into two parts; interval 1 (containing all data from 200 ms after the onset of the target up to 300 ms before the click response) and interval 2 (containing data from the last 300 ms before the click response)[13]. The selection for these time intervals was made because it would result in contrasting predictions with respect to co-activation. We predicted to find evidence for a critical temporal contingency in interval 1, but not in interval 2, given that participants would have finished evaluating the alternative responses and would fixate the picture they were going to click on while planning and executing their response (in the last 300 ms before the response). As a result, we expected the fit of model 6, that allows for different fixation patterns in different time intervals, to be better than the fit of model 1. The difference in fit of model 6 and model 1 was indeed significant $2\Lambda LL\ (14) = 87.49$, $p < .001$, which indicates that adding time as a covariate resulted in different transition matrices for the levels of time.

In model 7 we constrained the probability of switching from the English target to the Dutch competitor to be equal to the probability of switching from the English target to a distractor for the first time interval to test the specific hypothesis. When the fit of model 7 would be worse than the fit of model 6, our co-activation hypothesis would be confirmed for the first time interval. The difference in fit between model 6 and model 7 was significant, $2\Lambda LL\ (2) = 17.826$, $p < .0005$, showing that model 7 provided a worse fit to the data than model 6. The transition matrix belonging to model 6 showed a higher probability of switching from the English target to the Dutch competitor (.053) than to a distractor (.032) in time interval 1, thereby confirming the temporal contingency between the critical pictures in time interval 1. For time interval 2, the probability of a transition from the English target to a Dutch competitor (.014) was equal to that of a transition from the English target to a distractor (.014), showing that there are no temporal dependencies between the critical pictures in time interval 2, as hypothesized.

These results show that the time course within a trial can be divided into small time frames --as small as the researcher wishes to examine and the sampling rate of the eye tracking device allows-- thereby providing a more detailed look in the temporal aspect of language processing.

13 *It is possible to estimate continuous models for time, in which it is not necessary to define relevant time intervals a priori and one can assess the decay of the effects. Because we tested a priori hypotheses about time to validate our models, we assessed predefined time intervals.*

General discussion

In this article, we have described and demonstrated a method for assessing transitions (first-order shifts in overt attention) in an eye tracking paradigm. The main idea motivating this approach is that transitions between relevant visual referents provide an ideal measure for investigating momentary interactions between cognitive processes that are assumed to underlie viewing behavior. The crucial point is that when activation of one system affects activation of another system to the extent that an eye tracker can detect this activation (which will be discussed in the following paragraph), this should be observed in shifts in attention between their referents relative to other regions (with, e.g., more critical transitions in case of co-activation and less critical transitions in case of co-inhibition). As the two hypothetical examples in the introduction show, analyses on fixation proportions can erroneously suggest both the presence and absence of momentary interactions between cognitive processes. We argue that the stronger approach is to demonstrate interactions by directly assessing the eye movement transitions between critical regions within a trial. Even though we have illustrated this approach by assessing lexical competition in a visual world paradigm, its scope is much broader, encompassing all cases in which conclusions are drawn about momentary interplay between systems based on eye movement measures. The current approach of course still requires further considerations and development, which will be discussed in the following section. Next, we will embed this approach in a broader context by discussing applications in other theoretical frameworks and paradigms.

Future development and considerations

Even though transitions provide a more detailed window into interactions than proportions or delays, they are of course –like any other measure- not immune to false negatives. What if there are no transitions between critical fixations? It may be either the case that (a) there are no cognitive interactions or (b) there are cognitive interactions but they are too transient to be detected. In order to exclude the latter option that illustrates a false negative, at least the following assumptions have to be met. First, the duration of the interaction must be sufficiently long to result in concurrent eye movements. Second, the interaction must sufficiently increase the activation of both representations. If both representations are close to being equally probable then both referents may be fixated, but if one representation is much more strongly activated than the other this likely only results in fixations to the strongest referent. Third, the time needed to program and execute a saccade must not be longer than the temporal shifts in relative activation of both representations. Fourth, the concurrent fixation durations should be long enough to be detected but short enough to not conceal underlying shifts in representations. Fifth, the rate with which new bottom-up input is received is slow enough for competition effects to emerge in eye movements. As is the case for any method, these assumptions (and perhaps even more) implicitly underlie any eye tracking paradigm studying interactions. Given that critical transitions occurred in our experiment, these assumptions were evidently met in the current set-up. In the absence of critical transitions however, one would have to consider whether at least these conditions were met in order to be conclusive with respect to the absence of a cognitive interaction.

Another issue is that one should be careful when defining the time frames for assessing transitions. In the present example, time frames were set at 100 ms. However, the sampling rate of an eye tracker allows researchers to use smaller time frames (of up to 1 ms, depending on the eye tracking device). The reason for selecting larger time frames in our exemplary data lies in the relation between staying and switching probabilities. The time frames that are included in a markov model should reflect transition opportunities, i.e., the case in which it would have been possible to make an eye movement. The constraints of the human occulomotor system are such that individuals cannot make a saccade to a different picture every few milliseconds; it takes a minimum of 100 ms to plan and start an eye movement (Altmann, 2011) and an average of about 200-300 ms to plan and fully execute an eye movement until it lands on the target location. Therefore, if very small time frames had been used in the analyses, the results would have shown relatively higher staying probabilities and lower switching probabilities. In other words, when using smaller time frames, the staying probabilities would have been inflated relative to switching probabilities, possibly pushing switching probabilities toward floor. On the other hand, if larger time frames (of e.g., 300 ms) had been analyzed, transitions could be missed given that more than one transition might have occurred within a single time frame. Loss of transition information inherently results in a loss of power. By selecting time frames of 100 ms, we have attempted to find a balance between keeping the switching probabilities above floor without compromising the power of the test/measure. However, the time frames could also have been set at somewhat smaller or larger sizes (of e.g., 50 ms or 150 ms). Exactly what size of time frames is optimal when assessing transitions is open to examination.

An alternative to this time-based approach is to select an event-based approach, in which the staying probability is treated as the probability of making the next fixation on the same picture, and the switching probability is treated as the probability of making the next fixation on a different picture. In this case, the transition probabilities are calculated on event boundaries, rather than on time boundaries. However, it is still a matter of definition as to what would count as a fixation and what would not. Given that both approaches should use similar criteria for fixations (e.g., with a minimum duration of about 80 ms and within a certain region of interest), they would result in highly similar transition matrices.

When assessing transitions, one should further be alert for possible anticipatory baseline effects (cf., Barr, 2008; Barr, Gann, & Pierce, 2011; Dahan & Tanenhaus, 2005; Dahan, Tanenhaus, Salverda, 2007; McMurray, Tanenhaus, & Aslin, 2009; Wolter, Gorman, & Tanenhaus, 2011). One of the primary uses of visual world studies is to investigate issues related to information integration, in which constraining information is presented prior to presentation of the target word. If, for example, the preceding context in our exemplary study had biased towards the Dutch depiction, the probability of fixating these critical pictures would have been elevated. Such an anticipatory baseline effect could yield a similar set of transitions as presented in the prediction matrices, thereby confounding the present findings in favor of cross-language activation. However, no anticipatory baseline effects were expected to be present in the current study. Item selection was handled with great care and no constraining context was presented prior to the target word. Nevertheless, possible anticipatory baseline effects can be handled quite effectively. For example, by estimating a model over a baseline period before presentation of the target word, one can assess how transition probabilities after target presentation change with respect to this baseline. If the design of the experiment does not allow for such a baseline period, one might choose to eliminate trials in which the target or competitor were fixated at target onset (referred to as contingent analysis by e.g., Frank, Salverda, Jaeger, & Tanenhaus, 2009; Wolter, Gorman, & Tanenhaus, 2011) or might use transition probabilities from a control condition with identical visual displays as a baseline for transition probabilities in the critical condition (see Barr, Gann, & Pierce, 2011 for a discussion on handling anticipatory baseline effects).

Applications in other paradigms

Studying eye movement transitions need not be limited to visual world studies. This approach can be used in any eye tracking paradigm in which shifts in attention within a trial are of interest, such as scene perception, visual search, change blindness studies, or reading. The majority of studies on sentence and text reading, for example, often analyze averaged measures of forward or backward fixations as an index for comprehension, learning, and/or memory processes. For example, the proportion of regressions to previously fixated regions during sentence reading has been used as a measure for disrupted language processing, in which a higher proportion of regressions indicates greater processing difficulties (e.g., Frazier & Rayner, 1982). Furthermore, averaged regression durations on previous or following (parts of) sentences have been used for this purpose, with longer durations indicating greater processing difficulties (e.g., Just & Carpenter, 1987). Instead of assessing averaged regression durations or proportions across participants and trials, our current approach can provide estimations on the regression (i.e., transition) probabilities at an individual trial level. This enables researchers to distinguish different types of regression patterns across trials and/or individuals. By performing latent class markov modeling analyses, one could, for example, distinguish between latent regression patterns that are typical for skilled versus poor readers, or even unravel transition patterns between reading strategies within participants (see Simola, Salojärvi, & Kojo, 2008).

In general a reader is more likely to move forward in a sentence or text than to regress backward. Therefore, the successive locations of fixations (states) are not mutually independent. When such state dependencies are expected to occur in an eye tracking paradigm, predictions about viewing patterns should be formulated as a chain of transitions to different parts of the sentence or text, and should therefore include higher-order effects. With some minor adaptations to the syntax (see the online technical Appendix, URL, section B), higher-order hypotheses can easily be tested with the present markov modeling technique (see Althoff & Cohen, 1999, for an application of assessing second-order effects, or Simola, Salojärvi, & Kojo, 2008).

There are previous applications of the current approach in several fields. For example, Simola, Salojärvi, and Kojo (2008) distinguished three latent states (classes) of viewing behavior in a reading paradigm, referred to as scanning, reading, and decision making. In this study, each participant performed three types of information search tasks on a list of titles: simple word search, a question–answer task, and a task in which they had to select the subjectively most interesting topic. The authors found that the different tasks were characterized by different transition patterns between states, suggesting that readers switch between processing strategies (states), and that they do this in a different order depending on the information search task at hand. Other researchers have used this approach to distinguish local (reflecting the extraction of detailed information) from global (reflecting the redirection of attention) states of visual attention when people study advertisements in magazines (Liechty, Pieters, & Wedel, 2003). Furthermore, Althoff and Cohen (1999) and Henderson, Falk, Minut, Dyer, and Mahadevan (2000) have used markov model analyses to study sequences of fixations (referred to as 'scan patterns') that accompany face perception and recognition. Both studies found dependencies in successive fixation positions when perceiving an unknown face. For example, when fixating one of the two eye locations of the face, it was more likely that the successive fixation was directed to the other eye than to any other location in the face (Henderson et al., 2000).

These examples not only show that transition patterns can be used in different eye tracking paradigms, but also within different theoretical frameworks[14]. One could envisage numerous types of research domains that would benefit from assessing transitions in a fixation pattern, such as studies assessing attentional biases, approach-avoidance mechanisms, object permanence in young children, or social interactions.

Conclusion

Whenever the aim is to study a momentary interaction between cognitive systems, it is not sufficient to conclude that the visual referents of both systems were fixated across trials and participants. Rather, assessing whether a fixation to the referent of one system was temporally dependent on a fixation to the referent of the other system *can* provide a good indication. Analyzing eye movement transitions using a multilevel markov approach is an ideal way to tap into such temporal dependencies. In this paper, we have demonstrated that analyzing transitions cannot only provide useful, but also essential, insights in human viewing behavior and its underlying cognitive processes.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable comments on an earlier version of the manuscript. We further thank the members of the Language and Cognition Group and the Memory Lab at the Erasmus University Rotterdam for their feedback. Finally, we thank Mark Adriaans, Marcel Boom, and Christiaan Tieman of the Erasmus Behavioral Lab for their help with the data.

14 On an even higher level, the multilevel markov modeling approach is not even restricted to eye tracking data. It could, for example, be used to assess transitions in any type of responses (e.g., behavioral or ERP) on different trials within a block, different blocks within a task, different tasks within a participant, et cetera.

Online technical Appendix

A. Data structure

Table A.1 shows an exemplary data structure as used for multilevel markov modeling in the software package Latent Gold 4.5. Each single time frame *t* within a participant (variable 'pp') and within a trial (variable 'item') is represented by one record. In the present data, the total number of items was 14, the total number of participants was 35. The total number of time frames varied between trials (depending on the response latency), but the maximum number of time frames was fixed at 35 (corresponding to 3500 ms after target onset). Time interval was based on the time frames; the final 3 time frames corresponded to time interval 2 and the preceding time frames corresponded to time interval 1. The dependent variable 'picture' had four nominal categories (target [1], competitor [2], distractor one [3], and distractor two [4]) and varied at the level of time frames.

Table A.1. Exemplary data structure for markov modeling in Latent Gold.

Pp	item	time frame	picture	accuracy	time interval
1	1	3	1	1	1
1	1	4	.	1	1
1	1	5	2	1	1
1	1	..	..	1	..
1	1	..	..	1	..
1	1	T	1	1	2
1	2	3	2	1	1
1	2	4	2	1	1
1	2	..	..	1	..
1	2	T	3	1	2
1	3	3	2	0	1
1	3	..	..	0	..
1	3	T	1	0	2
1	..	..	..	..	..
1	..	..	..	..	..
1	K	3	2	1	1
1	K	4	2	1	1
1	K	..	..	1	..
1	K	T	1	1	2
2	1	3	2	1	1
2	1	T	2	1	2
2	2	1	4	1	1
2	..	..	..	..	..
2	K	T	2	1	2
3	1	T	1	1	2
3	..	T	3	..	2
3	K	T	2	1	2
..	..	..	..	..	..
..	..	..	..	..	..
N	1	T	1	1	2
N	K	T	1	1	2

B. Syntax

The syntax presented below presents the code used for multilevel markov modeling in Latent Gold. Explanation of the preceding syntax is presented in italics. A more detailed explanation of Latent Gold syntax can be found in the syntax guide (Vermunt & Magidson, 2008) and technical guide (Vermunt & Magidson, 2005) of Latent Gold 4.0 which can be downloaded at :
http://www.statisticalinnovations.com/products/LGsyntax_Manual.pdf and
http://www.statisticalinnovations.com/products/LGtechnical.pdf

options

algorithm

tolerance=1e-008 emtolerance=0,01 emiterations=250 nriterations=100;

// To find maximum likelihood estimates for the model parameters, Latent GOLD uses both the EM (expectation maximization) and the NR (Newton-Raphson) algorithm. The estimation process starts with a number of EM iterations. When close enough to the final solution, the program switches to Newton-Raphson. This is a way to exploit the advantages of both algorithms; that is, the stability of EM even when it is far away from the optimum and the speed of Newton-Raphson when it is close to the optimum. In our analysis, the number of EM iterations was set to 250, and the number of Newton-Raphson iterations was set to 100. The converge criterion for the EM iterations (tolerance) was set at .01 and for the Newton-Raphson method it was set to .00000001. As soon as the difference in loglikelihood was below this converge criterion, the estimation process finished.

startvalues

seed=0 sets=10 tolerance=1e-005 iterations=50;

// Latent GOLD generates random start values for the estimation process. These differ every time that a model is estimated because the seed of the random number generator is obtained from the system time, as long as the technical option 'seed' equals 0. In our analysis we used 10 sets of maximally 50 iterations to search for the best start values.

missing includeall;

// In markov modeling all records with missing values should be retained. The reason for this is that eliminating records with missing values would distort the structure of the data file.

output

parameters=effect standarderrors estimatedvalues;

// The output shows the parameter estimates (using effect coding), standard errors, and the estimated values (staying and switching probabilities).

variables

groupid pp;

//The participants form the highest level of the multilevel model.

caseid item;

//Items are nested within participants.

dependent picture nominal;

//The four pictures form a nominal dependent variable.

independent accuracy nominal coding=first, time nominal coding=last;

//The four pictures form a nominal dependent variable.

latent

MClass nominal dynamic 4 coding = first,

// Mclass is a time-specific four-class (dynamic) latent variable. In fact, this syntax defines a hidden markov model in which the states are a latent class variable (Mclass) which may contain measurement error. Because we used a manifest markov model in which the states were observed we added the constraint z=-100 (see below) which changes the hidden markov model into a manifest markov model without measurement error.

Dummy coding is used with the first class as a reference category.

Ifactor continuous group;

Kfactor continuous group;

Gfactor continuous group;

// Kfactor is the random effect that is used to model the dependency of items within a participant on time t=1. IFactor is the random effect that is used to model the dependency of items within a participant from time t-1 to time t. Gfactor is the random effect of the covariate to model the dependency of items within a participant from time t-1 to time t.

equations model 0

(1) IFactor;

(1) KFactor;

// The parameter of the continuous latent variables is fixed to 1 in order to identify the model.

MClass[=1] <- (a)1 + (b)KFactor

// This equation is the logit equation for the initial state probabilities at t=1 for the four pictures. This equation contains only a fixed intercept.

MClass <- (c)1|MClass[-1] + (d) Ifactor|MClass[-1];

// This is the logit equation for the transition probabilities between t-1 and t. The dependent variable MClass is the picture a participant fixates at time t.

picture<- (z~err) 1 | MClass;

// The equation links the state (hidden markov) at time t to the dependent variable (picture) at time point t.

z=-100;

// A simple markov model without measurement error is obtained by expanding the specification with the restriction z=-100. When this constraint is omitted a hidden markov model is estimated.

equations model 1

(1) IFactor;

(1) KFactor;

// The parameter of the continuous latent variables is fixed to 1 in order to identify the model.

MClass[=1] <- (a)1 + (b)KFactor

// This equation is the logit equation for the initial state probabilities at t=1 for the four pictures. This equation contains only a fixed intercept.

MClass <- (c)1|MClass[-1] + (d) Ifactor|MClass[-1];

// This is the logit equation for the transition probabilities between t-1 and t. The dependent variable MClass is the picture a participant fixates at time t.

picture<- (z~err) 1 | MClass;

//The equation links the state (hidden markov) at time t to the dependent variable (picture) at time point t.

z=-100;

// A simple markov model without measurement error is obtained by expanding the specification with the restriction z=-100. When this constraint is omitted a hidden markov model is estimated.

a[1,3]=c[1,2];

b[1,3]=b[1,2];

c[1,3]=c[1,2];
c[4,2]=c[3,3];
c[2,2]=c[2,3];
c[4,3]=c[3,2];
c[4,1]=c[3,1];

d[1,3]=d[1,2];
d[4,2]=d[3,3];
d[2,2]=d[2,3];
d[4,3]=d[3,2];
d[4,1]=d[3,1];

// The restrictions defined above are equality restrictions which make the initial probabilities and transition probabilities for the two distractor pictures the same.

restrictions model 2 in addition to the restrictions of model 1

a[1,1]=a[1,3];

b[1,1]=b[1,3];

c[1,2]=c[1,1];
c[3,1]=c[4,2];
c[2,2]=c[3,1];
c[3,2]=c[2,1];

d[1,2]=d[1,1];
d[3,1]=d[4,2];
d[2,2]=d[3,1];
d[3,2]=d[2,1];

// The restrictions defined above make the initial probabilities and transition probabilities for the Dutch competitor picture and the two distractor pictures the same.

restrictions model 3 in addition to the restrictions of model 1

```
c[1,2]=c[1,1];
d[1,2]=d[1,1];
```

// The restrictions make the transition probability from the English target picture to the Dutch competitor picture equal to the probability from the English target picture to the distractor picture.

equations model 4

```
    (1) IFactor;
    (1) KFactor;
    (1) GFactor;
```

// The parameter of the continuous latent variables is fixed to 1 in order to identify the model.

```
    MClass[=1] <- (a)1 + (b)KFactor + (e)GFactor Accuracy
    MClass <-  (c)1|MClass[-1] + (d) Ifactor|MClass[-1]+(f)Accuracy|MClass[-1] +(g)
    GFactor Accuracy|MClass[-1];;picture<- (z~err) 1 | MClass;
        z=-100;

a[1,3]=c[1,2];
b[1,3]=b[1,2];
e[1,3]=e[1,2];

c[1,3]=c[1,2];
c[4,2]=c[3,3];
c[2,2]=c[2,3];
c[4,3]=c[3,2];
c[4,1]=c[3,1];

d[1,3]=d[1,2];
d[4,2]=d[3,3];
d[2,2]=d[2,3];
d[4,3]=d[3,2];
```

```
d[4,1]=d[3,1];

f[1,3]=f[1,2];
f[4,2]=f[3,3];
f[2,2]=f[2,3];
f[4,3]=f[3,2];
f[4,1]=f[3,1];

g[1,3]=g[1,2];
g[4,2]=g[3,3];
g[2,2]=g[2,3];
g[4,3]=g[3,2];
g[4,1]=g[3,1];
```

// The restrictions defined above are equality restrictions which make the initial probabilities and transition probabilities for the two distractor pictures the same.

restrictions model 5 in addition to the restrictions of model 4

```
c[1,2]=c[1,1];
d[1,2]=d[1,1];
f[1,2]=f[1,1];
g[1,2]=g[1,1];
```

// The restrictions make the transition probability from the English target picture to the Dutch competitor picture equal to the probability from the English target picture to the distractor picture.

equations model 6

```
    (1) IFactor;
    (1) KFactor;
    (1) GFactor;

    MClass[=1] <- (a)1 + (b)KFactor
    MClass <-  (c)1|MClass[-1] + (d) Ifactor|MClass[-1]+(f)time |MClass[-1] +(g)
    GFactor time |MClass[-1];;picture<- (z~err) 1 | MClass;
        z=-100;
```

a[1,3]=c[1,2];

b[1,3]=b[1,2];
c[1,3]=c[1,2];
c[4,2]=c[3,3];
c[2,2]=c[2,3];
c[4,3]=c[3,2];
c[4,1]=c[3,1];

d[1,3]=d[1,2];
d[4,2]=d[3,3];
d[2,2]=d[2,3];
d[4,3]=d[3,2];
d[4,1]=d[3,1];

f[1,3]=f[1,2];
f[4,2]=f[3,3];
f[2,2]=f[2,3];
f[4,3]=f[3,2];
f[4,1]=f[3,1];

g[1,3]=g[1,2];
g[4,2]=g[3,3];
g[2,2]=g[2,3];
g[4,3]=g[3,2];
g[4,1]=g[3,1];

restrictions model 7 in addition to the restrictions of model 6

c[1,2]=c[1,1];
d[1,2]=d[1,1];
f[1,2]=f[1,1];
g[1,2]=g[1,1];

// The restrictions make the transition probability from the English target picture to the Dutch competitor picture equal to the probability from the English target picture to the distractor picture.

C. Restrictions of the multilevel markov models

Model 1. Distractor probabilities are equal

For each trial, staying probabilities and switching probabilities to the two distractors were assumed to be equal because assignment of distractor pictures to either of two conditions was random, with no presupposed differences between the two pictures. Therefore, we expected that both the initial probabilities and the transition probabilities (both the staying and switching probabilities) were equal for these two pictures. In order to test this restriction we put the following restrictions on the initial fixation parameters:
$\alpha_{s=dis1,t=1} = \alpha_{s=dis2,t=1}$ and $g_{s=dis1,t=1} = g_{s=dis2,t=1}$

We put restrictions on the intercept transition parameters:

$$\gamma_{0,v=dis1|u=en} = \gamma_{0,v=dis2|u=en}$$
$$\gamma_{0,v=dis1|u=du} = \gamma_{0,v=dis2|u=du}$$
$$\gamma_{0,v=dis1|u=dis1} = \gamma_{0,v=dis2|u=dis2}$$
$$\gamma_{0,v=dis2|u=dis1} = \gamma_{0,v=dis1|u=dis2}$$
$$\gamma_{0,v=du|u=dis1} = \gamma_{0,v=du|u=dis2}$$
$$\gamma_{0,v=en|u=dis1} = \gamma_{0,v=en|u=dis2}$$

and the same restrictions on the random effect:

$$r_{v=dis1|u=en} = r_{v=dis2|u=en}$$
$$r_{v=dis1|u=du} = r_{v=dis2|u=du}$$
$$r_{v=dis1|u=dis1} = r_{v=dis2|u=dis2}$$
$$r_{v=dis2|u=dis1} = r_{v=dis1|u=dis2}$$
$$r_{v=du|u=dis1} = r_{v=du|u=dis2}$$
$$r_{v=en|u=dis1} = r_{v=en|u=dis2}$$ [15]

Model 2. Dutch and distractor probabilities are equal

If languages are not co-activated, we would expect the probabilities of fixating the critical Dutch depiction and the distractor depictions to be equal. In addition to the restrictions above, we put the following constraint on the initial parameters:
$\alpha_{s=du,t=1} = \alpha_{dis1,t=1}$ and $g_{s=du,t=1} = g_{s=dis1,t=1}$

15 *Note that this parameter is redundant since the parameters sum up to zero.*

We put additional restrictions on the intercept transition parameters:

$\gamma_{0,v=du|u=en}=\gamma_{0,v=dis1|u=en}$
$\gamma_{0,v=du|u=du}=\gamma_{0,v=dis1|u=dis1}$
$\gamma_{0,v=en|u=du}=\gamma_{0,v=en|u=dis1}$
$\gamma_{0,v=du|u=dis1}=\gamma_{0,v=dis1|u=du}$

and the same additional restrictions on the random effect:

$r_{v=du|u=en} = r_{v=dis1|u=en}$
$r_{v=du|u=du} = r_{v=dis1|u=dis1}$
$r_{v=en|u=du} = r_{v=en|u=dis1}$
$r_{v=du|u=dis1} = r_{v=dis1|u=du}$

Model 3. The switching probability from English to Dutch is equal to the switching probability from English to a distractor.

The crucial hypothesis was that the probability of switching from the English target to the Dutch competitor was higher than switching from the English target to a distractor, because this would demonstrate a temporal contingency between fixations on the English and Dutch pictures. This hypothesis can be tested by adding the following constraint to the transition parameters of model 1:

$\gamma_{0,v=du|u=en}= \gamma_{0,v=dis1|u=en}$
$r_{v=du|u=en} = r_{v=dis1|u=en}$

Model 4. Accuracy is added to model 1 as a covariate.

This hypothesis can be tested by adding the following constraint to the constraints that were already formulated in model 1:

$\alpha_{l,s=du,t=1}= \alpha_{l,s=dis1,t=1}$
$\gamma_{l,v=du|u=en}= \gamma_{l,v=dis1|u=en}$
$\gamma_{l,v=du|u=du}= \gamma_{l,v=dis1|u=dis1}$
$\gamma_{l,v=en|u=du}= \gamma_{l,v=en|u=dis1}$
$\gamma_{l,v=du|u=dis1}= \gamma_{l,v=dis1|u=du}$

and the same additional restrictions on the random effect of the covariate accuracy:

$r_{l,v=du|u=en} = r_{l,v=dis1|u=en}$
$r_{l,v=du|u=du} = r_{l,v=dis1|u=dis1}$
$r_{l,v=en|u=du} = r_{l,v=en|u=dis1}$
$r_{l,v=du|u=dis1} = r_{l,v=dis1|u=du}$

Model 5. The switching probability from English to Dutch is equal to the switching probability from English to a distractor, for the accurate clicks.

The following restrictions are added to model 4:

$\gamma_{l,v=du|u=en} = \gamma_{l,v=dis1|u=en}$

$r_{l,v=du|u=en} = r_{l,v=dis1|u=en}$

Model 6. Time interval is added to model 1 as a covariate.

This hypothesis can be tested by adding the following constraint to the constraints that were already formulated in model 1:

$\alpha_{l,s=du,t=1} = \alpha_{l,s=dis1,t=1}$

$\gamma_{l,v=du|u=en} = \gamma_{l,v=dis1|u=en}$

$\gamma_{l,v=du|u=du} = \gamma_{l,v=dis1|u=dis1}$

$\gamma_{l,v=en|u=du} = \gamma_{l,v=en|u=dis1}$

$\gamma_{l,v=du|u=dis1} = \gamma_{l,v=dis1|u=du}$

and the same additional restrictions on the random effect of the covariate time:

$r_{l,v=du|u=en} = r_{l,v=dis1|u=en}$

$r_{l,v=du|u=du} = r_{l,v=dis1|u=dis1}$

$r_{l,v=en|u=du} = r_{l,v=en|u=dis1}$

$r_{l,v=du|u=dis1} = r_{l,v=dis1|u=du}$

Model 7. The switching probability from English to Dutch is equal to the switching probability from English to a distractor, for the first time interval.

The following restrictions are added to model 6:

$\gamma_{l,v=du|u=en} = \gamma_{l,v=dis1|u=en}$

$r_{l,v=du|u=en} = r_{l,v=dis1|u=en}$

Chapter 4

Out of mind, out of sight: Language affects perceptual vividness in memory

Abstract

We examined whether language affects the strength of a visual representation in memory. Participants studied a picture, read a story about the depicted object, and then selected out of two pictures the one whose transparency level most resembled that of the previously presented picture. The stories contained two linguistic manipulations that have been demonstrated to affect concept availability in memory, i.e., object presence and goal-relevance. The results show that described absence of an object caused people to select the most transparent picture more often than described presence of the object. This effect was not moderated by goal-relevance, suggesting that our paradigm tapped into the perceptual quality of representations rather than, for example, their linguistic availability. We discuss the implications of these findings within a framework of grounded cognition.

This chapter is published as: Vandeberg, L., Eerland, A., & Zwaan, R.A. (2012). Out of Sight, Out of Mind: Language Affects Perceptual Vividness in Memory. *PLOS ONE*, 7(4): e36154. doi:10.1371/journal.pone.0036154.

Introduction

When we comprehend language, we create a mental representation of the situation that is described by the text (Bower & Morrow, 1990; Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Embodied theories of language propose that the information in such situation models is not merely abstract or symbolic, but rather is grounded in our bodily experiences (Barsalou, 1999; Pulvermüller, 1999; Zwaan, 2003). This proposal implies that the representations we create or retrieve during language processing are similar to the representations we created when we actually experienced the respective situation. As a result, these conceptual representations include perceptual, lexical, semantic, and functional features and are shaped by people's world knowledge and beliefs (Murphy & Medin, 1985). There is ample evidence that people retrieve perceptual (Fincher-Kiefer, 2001; Horton & Rapp, 2003; Lupyan & Thompson-Schill, 2011) and motoric (Glenberg & Kaschak, 2002; Tucker & Ellis, 2004) information from memory while processing language. For example, they represent the visual features of a described object or action (such as shape, orientation, color, size, or direction of motion) while reading about a situation (Borghi & Riggio, 2009; Dils & Boroditsky, 2010; Hirschfeld & Zwitserlood, 2011; Holt & Beilock, 2006; Kaschak et al., 2005; Lincoln, Long, & Baynes, 2007; Stanfield & Zwaan, 2001; Zwaan, Madden, Yaxley, & Aveyard, 2004; Zwaan, Stanfield, & Yaxley, 2002).

We ask whether language can affect the activation of a concept to such an extent that it improves or reduces the quality of a visual representation. Only a few studies that we know of have explored the quality of visual mental representations by manipulating the resolution (Yaxley & Zwaan, 2007), realism (Holmes & Wolff, 2011), vividness (Risen & Critcher, 2011) or spatial frequency (Hirschfeld & Zwitserlood, 2011) of visual stimuli. In the current study we refer to a representation's quality in terms of *perceptual vividness.* We isolate the perceptual component of a representation in our definition of vividness, however, perceptual information is not exhaustive of a conceptual representation (Lupyan & Thompson-Schill, 2011). We make the critical assumption that when the perceptual component of a representation is activated to a greater extent, the availability of visual information -such as outline or color- will increase and thereby also what we call the vividness of a mental representation in memory. We assessed whether (a) the general activation of a concept, or (b) merely the activation of a concept's visual component affects vividness of a mental representation.

Many types of linguistic information can have an impact on the availability of concept information in memory. For example, objects that are present (MacDonald & Just, 1989), visible (Horton & Rapp, 2003), or spatially close (Glenberg, Meyer, & Lindem, 1987) to a protagonist in a described situation are more accessible than objects that are absent, occluded, or farther away, respectively. For example, in one study participants read short stories describing a character's view of an object that was either blocked (e.g., by a curtain) or not (Horton & Rapp, 2003). They were slower to respond to verification questions about objects that were occluded from view than to visible objects, suggesting that the accessibility of objects is reduced during recall when these objects are absent from the protagonist's view in a described situation. Another factor that affects the accessibility of an object is whether that object is goal-relevant. Information that is relevant to the protagonist's goal is retrieved faster than irrelevant information (Albrecht & Myers, 1995; Dopkins, Klin, & Myers, 1993; Lutz & Radvansky, 1997; Magliano, Taylor, & Kim, 2005; Suh & Trabasso, 1993; Zwaan & Radvansky, 1998). Responses to probe words are faster after reading texts in which a goal was achieved, compared to control texts that describe a simple completed action (Dopkins, Klin, & Myers, 1993; Lutz & Radvansky, 1997). This suggests that goal category information is more accessible than information that is irrelevant (neutral) to the goal.

To investigate whether and how concept availability affects the perceptual vividness of the associated conceptual representation in memory, we created short stories in which we manipulated both object presence (versus absence) and goal-relevance (versus irrelevance). In these described situations, object presence alters the visual components of a conceptual representation, whereas goal-relevance may alter the conceptual representation in a non-perceptual way. We formulated two main hypotheses.

The first major hypothesis is that increased availability of a concept in memory leads to increased perceptual vividness of the accompanying representation. Thus, both object presence and goal-relevance should exert an effect on recalled vividness. The most straightforward evidence would come from a pattern in which both object presence and goal-relevance increase perceptual vividness of the concept in memory. However, it is also possible that object presence and goal-relevance interact in specific ways. Out of the space of possible hypotheses in support of a mapping between concept availability and vividness, the following two seem particularly relevant.

Research on goal-relevance has shown a greater availability for objects that are relevant to the protagonist's goals rather than irrelevant (Albrecht & Myers, 1995; Dopkins, Klin, & Myers, 1993; Lutz & Radvansky, 1997; Magliano, Taylor, & Kim, 2005; Suh & Trabasso, 1993; Zwaan & Radvansky, 1998). As a result, presence or absence of an object may only be sufficiently salient for relevant objects. In other words, information about presence or absence of objects may differentiate the availability of relevant concepts but not of irrelevant concepts (Lutz & Radvansky, 1997). In this case, we would expect an interaction in which an increased vividness for present compared to absent objects would only be detected for relevant objects. (hypothesis 1a).

However, there is a parallel between our manipulation of presence and manipulations of negation in the literature. There is evidence that negated concepts do not simply reduce accessibility of a concept, but are even replaced by alternative opposites later in processing (Ferguson, Sanford, & Leuthold, 2008; Giora, 2006; Hasson & Glucksberg, 2006; Kaup, Lüdtke, & Zwaan, 2006). These robust effects of negation may result in a floor level accessibility to absent concepts, regardless of their goal relevance. In this case, we would expect an interaction in which an increased vividness for relevant compared to irrelevant objects would only be detected for present objects. (hypothesis 1b).

The major alternative hypothesis is that only object presence affects perceptual vividness. Describing that an object is or is not present involves the visual aspects of a situation, whereas describing that an object is or is not relevant to a protagonist's goals does not. After all, if objects are present they are visible regardless of whether they are relevant to a protagonist's goals. Thus, this hypothesis predicts a main effect of presence but no main effect of goal relevance and no interaction. (hypothesis 2).

Experiment 1

Method

Ethics Statement

All participants were recruited online and voluntarily subscribed for participation in all of the described experiments. Written consent was not obtained because the experiment was noninvasive. This is in accordance with departmental practice approved by the Ethics Committee of Psychology (ECP) at the Erasmus University Rotterdam, the Netherlands.

Participants

242 participants were recruited online through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 35 (SD=12) and contained 153 females (63%). 225 participants (93%) reported being a native speaker of English[16]. All participants were residents of the USA and were compensated with $1.50 for their participation, which required approximately 25 minutes.

16 *These demographics are based on 241 participants, because the demographic data of one participant were missing. This participant and the 16 participants that did not report English as their native language were not excluded from the sample, because inclusion did not alter the result pattern.*

Materials and design

Two versions of twenty critical short stories were created (see online Appendix A). Each story consisted of five sentences. The first three sentences of each story introduced an event and included information about the characters, their actions, and a location (see example story 1). The fourth sentence introduced a critical object that was either absent or present in the described situation, which was defined by placing the articles 'a' or 'no' before the object noun. The object was relevant to the protagonist's goal in one version of the story (version a) but irrelevant to the protagonist's goal in the other version (b). Goal-relevance was foregrounded in the first three sentences; the fourth sentence was similar for both versions of the stories. The fifth and final sentence did not mention the object and was neutral with respect to the presence or goal-relevance of the object.

1. Jennifer was on a mountain bike trip with her friends. *(1)*
2a. After an hour, everyone had gotten thirsty.
2b. The terrain was hilly and the trails were difficult.
3. (After an hour,) They stopped to take a break at a picnic area.
4. Jennifer saw a/no water fountain there.
5. After fifteen minutes, the group climbed back on their bikes.

Each story was paired with a picture that corresponded to the described object. Additionally, 20 filler picture-story pairs were created in which the matched picture did not correspond to the described object. The transparency level of all 40 pictures was adjusted to three different values using Adobe® Photoshop® software. A 50% transparency level served as a baseline condition, whereas a 45% transparency level reflected a slightly more transparent (less vivid) version of the picture and a 60% transparency level was slightly more opaque (more vivid)[17], see online Appendix B.

This approach of mapping visual transparency levels onto representational vividness is inspired by work in social psychology (Risen & Critcher, 2011). In that study, people set the recalled transparency of previously processed pictures of a hot desert to less transparent (more opaque) when seated in a hot room than when seated in a cold room. The researchers interpreted this finding as evidence for perceptual fluency, in the sense that participants who experienced the visceral state of warmth, constructed more vivid and fluent mental representations of hot (versus cold) images. In the current study, we assessed how both *linguistic* factors affected recall of the transparency level of previously presented pictures.

17 In a pilot study with identical absolute differences between pictures (40%, 50%, 60%), people showed a tendency to select more opaque pictures over more transparent pictures when matching them to baseline pictures. In an attempt to select more comparable relative differences in transparency, we created a smaller absolute difference for the more transparent pictures than for the more opaque pictures.

We created two lists, one for goal-relevant and one for goal-irrelevant stories, thereby using goal-relevance as a between-subjects factor. In each list, half of the stories described the presence of an object and the other half described the absence of an object, thereby using object presence as a within-subjects factor. The lists were counterbalanced across subjects and the picture-story pairs were presented in randomized order within subjects.

Procedure

The experiment was programmed and presented in the Qualtrics survey research suite (http://www.qualtrics.com). Participants were instructed to (1) look at the presented picture, (2) read a short story, (3) decide which of two presented pictures best matched the picture they had seen previously, and (4) answer a question about the story. Each trial started with the presentation of a single picture. Critical pictures were presented at a 50% transparency level, whereas half of the filler pictures were presented at 45% and the other half at 60%. When participants clicked on a button on the screen, the picture disappeared and the story appeared. After reading the story and clicking a button, the 45% version of the previously seen picture appeared on the left of the screen and the 60% version appeared on the right. Participants indicated which version best matched the picture they had seen in the first part of the trial by checking the corresponding box. Additionally, a comprehension question that required a yes/no response followed one fourth of the trials to make sure that participants read the stories properly. After participants had answered the question, the next trial started. After completing all 40 trials, participants answered 20 questions about the absence or presence of the critical objects by checking the 'yes' or 'no' response box (e.g., 'In the story about the mountain bike trip, did Jennifer see a water fountain in the picnic area?'). Finally, participants filled out several demographical questions.

Results

Accuracy on the comprehension questions (M=0.85; SD=0.14) and the final questions about presence of the objects (M=0.84; SD=0.13) was high and above chance level (t (241)= 38.79, p <.0001 and t (241)= 39.14, p < .0001, respectively), indicating that participants properly read the stories. We calculated the average hit rate for both transparency levels (45% or 60%) across items for each participant, see Figure 1. Because the two hit rates are complementary, we only discuss the hit rates to the most transparent picture. A 2 (presence) x 2 (goal-relevance) x 2 (list) mixed design ANOVA revealed only a significant main effect of presence (F (1,238)= 10.69, p=.001, η^2=.043)[18]. Participants selected the most transparent picture more often when reading about an absent object (in 71.7% of the trials) than when reading about a present object (in 66.4% of the trials). Importantly, no other effects were significant, showing that goal-relevance did not significantly affect participants' decisions (main effect of goal-relevance: F (1,238) = 0.50, p = .48, η^2= .00; interaction goal-relevance*presence: F (1,238) = 0.64, p = .43, η^2= .00).

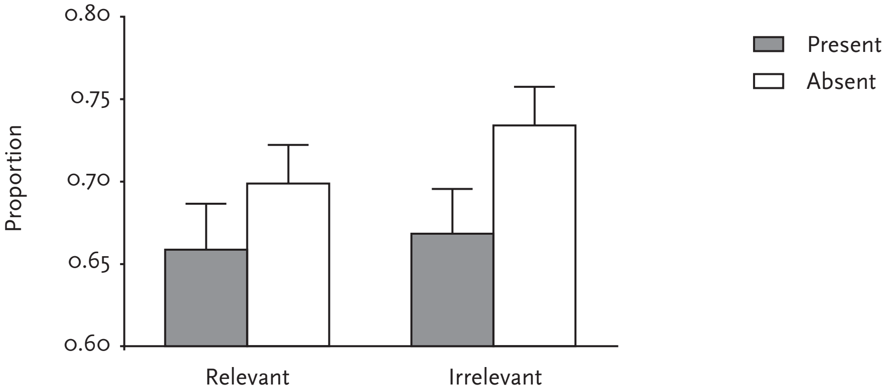


Figure 1. Proportion of hits on the most transparent picture per category in Experiment 1. Object presence represents a within-subjects factor and goal-relevance represents a between-subjects factor. Error bars represent the standard errors of the means.

Experiment 2

The results from Experiment 1 showed that participants selected the transparent picture more often when reading about an absent object than a present object. This suggests that the recalled vividness of the picture was reduced after reading about absence of the referent object in the described situation (relative to the object being present in the described situation). These findings demonstrate that vividness (quality) indeed is an aspect of perceptual representations, and that this aspect is affected by language. Furthermore, goal-relevance did not significantly affect participants' decisions. This finding supports our second main hypothesis that a change in availability of a concept in memory does not necessarily lead to an altered perceptual vividness of the accompanying representation. This suggests that the current paradigm isolates the perceptual quality of a concept rather than the overall concept availability in a more abstract way

18 List was included as a between-subjects factor. Effects for the list variable are not reported, given the lack of theoretical relevance (Pollatsek & Well, 1995).

(the way a probe word might do). In order to (a) optimize the possibility of detecting an effect of goal-relevance by including it as a within-subjects factor, (b) assess whether participants were aware of the purpose of the experiment, and (c) replicate our findings, we performed a follow-up experiment.

Method

Participants

229 participants were recruited online through Amazon's Mechanical Turk (http://www.mturk.com). This sample had a mean age of 32 (SD=11) and contained 146 females (64%). 220 participants (96%) reported being a native speaker of English. They were all residents of the USA and were compensated with $1.50 for their participation, which required approximately 25 minutes.

Materials and design

The materials and design were identical to those of Experiment 1, except that goal-relevance was treated as a within-subject factor. This resulted in 10 goal-relevant and 10 goal-irrelevant stories within a list, in half of which the object was present and half of which it was absent for each condition (5 relevant-present, 5 relevant-absent, 5 irrelevant-present, 5 irrelevant-absent). Four lists were created that were counterbalanced across subjects, the picture-story pairs were presented in randomized order within subjects.

Procedure

The procedure was identical to that of Experiment 1 except that participants were prompted about the purpose of the experiment. The prompt was presented after all the picture-story-picture trials but before the presence questions and the demographic questions. This prevented the participants from post-rationalizing the purpose of the study after focusing on absent versus present objects in the presence questions. They received the following instruction: 'We would like to know what you think this survey is about. In the space below, please take your best guess at describing the purpose of this study'.

Results

Accuracy on the comprehension questions (M=0.85; SD=0.15) and the final questions about presence of the objects (M=0.83; SD=0.14) was high and above chance level (t (225)= 36.02, $p < .0001$ and t (225)= 36.21, $p < .0001$, respectively), indicating that participants had properly read the stories. Four participants were excluded from the sample because, when prompted, they associated the presence of the objects in the stories with the selection of different transparency levels of the pictures. We calculated the average hit rate for each of both transparency levels across items for each participant (see Figure 2) and will only discuss the hit rates for the most transparent picture. Consistent with our findings from Experiment 1, a 2 (presence) x 2 (goal-relevance) x 4 (list) mixed design ANOVA revealed only a significant main effect of presence (F $(1,221) = 5.42$, $p=.02$, $\eta^2=.024$)[18]; the transparent picture was selected in 69% of the trials with an absent object and 65.5% of the trials with a present object. Importantly, no other effects were significant, showing that goal-relevance did not affect participants' decisions (main effect of goal-relevance: F $(1,221) = 1.14$, $p = .29$, $\eta^2= .01$; interaction goal-relevance*presence: F $(1,221) = 0.22$, $p = .64$, $\eta^2= .00$).

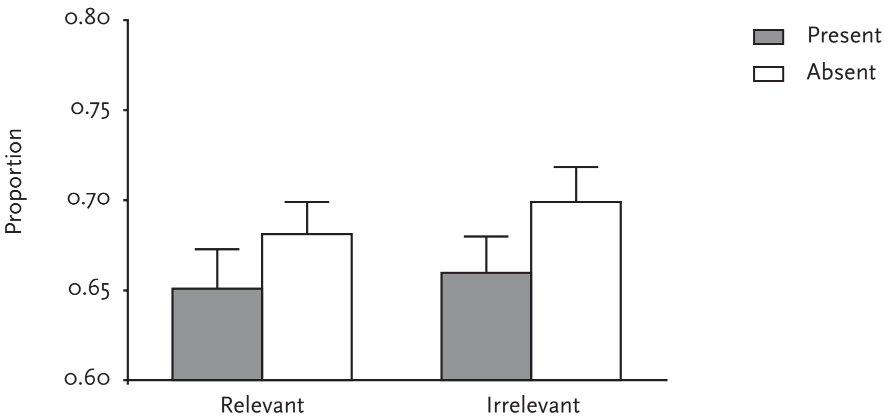


Figure 2. Proportion of hits on the most transparent picture per category in Experiment 2. Object presence and goal-relevance represent within-subjects factors. Error bars represent the standard errors of the means.

Experiment 3

The results of Experiment 2 demonstrate that transparency judgments were affected by presence of the target object in the referential situation, suggesting that object absence reduced the recalled vividness of an associated picture. Furthermore, the results again demonstrated that described goal-relevance did not affect transparency judgments. If the current transparency paradigm would provide a measure of concept availability per se, we would have expected goal-relevance to affect the results. Therefore, these findings suggest that our paradigm specifically taps into a perceptual aspect of conceptual representations.

A potential criticism to this conclusion could be that our manipulation of goal-relevance did not sufficiently distinguish between different levels of concept availability. If, for example, goal-relevance in our stories did not affect the availability of the stored concept, it would also not affect the perceived vividness of the concept. To ensure that our linguistic manipulations indeed resulted in distinct levels of concept availability (with a higher availability for present and relevant objects as opposed to absent and irrelevant objects), we performed a production experiment.

Methods

Participants

64 participants were recruited online through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 34 (SD=13) and contained 42 females (66%). 60 participants (94%) reported being a native speaker of English. All participants were residents of the USA and were compensated with $0.60 for their participation, which required 10-13 minutes on average.

Materials and Procedure

The stories were identical to those in Experiment 2 but the final sentence was removed. No filler stories were included in this experiment, because the only aspect that distinguished fillers from experimental stories in the previous experiments was whether the presented pictures matched the object in the story or not. Because Experiment 3 did not include pictures, the filler stories no longer served a purpose. We created four lists with twenty stories (5 relevant-present, 5 relevant-absent, 5 irrelevant-present, 5 irrelevant-absent) that were counterbalanced across subjects and randomized within subjects. Again, the experiment was programmed and presented in the Qualtrics survey research suite. Participants were instructed to create the fifth missing sentence that fitted the preceding four given sentences from the story.

We aimed to measure the participants' accessibility to the object at the time of sentence creation by analyzing the content of the created sentences. The rationale behind this is that a greater semantic relatedness between the described object and the created sentence would demonstrate an enhanced availability of the concept. For this purpose, we computed for each condition the semantic overlap between the critical object (e.g., water fountain) and the produced sentences (e.g., 'She went to the fountain and quenched her thirst.') by means of Latent Semantic Analysis (LSA: http://lsa.colorado.edu/). LSA is a technique for computing the similarity of text pairs by comparing their vector representations in a multi-dimensional vector space, which is created from a large text corpus. Higher similarity values for stories in which the object was present or goal-relevant would demonstrate that present and goal-relevant stories indeed resulted in increased concept availability as compared to stories in the absent or irrelevant condition.

Results

The similarity values computed by LSA ranged from -0.02 to 0.61. A 2 (object presence) * 2 (goal-relevance) * 4 (list)[18] mixed design ANOVA revealed a significant main effect of presence (F (1, 60) = 111.06, p < .0001, η^2=.65), showing that the produced sentences had greater semantic similarity to objects in the present rather than the absent condition (M=.29 and M=.16, respectively). Furthermore, a significant main effect of goal-relevance (F (1, 60)= 4.80, p <.05, η^2=.07) revealed a greater semantic overlap between produced sentences and goal-relevant objects (M=.23) as compared to irrelevant objects (M=.21). Finally, analyses showed a significant two-way interaction of object presence and goal-relevance (F (1, 60)= 4.09, p <.05, η^2=.06). Post-hoc paired-samples t-test revealed that the effect of goal-relevance only occurred for present objects ($M_{diff\_present}$ = 0.311 – 0.267 = 0.045, t (63)=2.53, p < .05) but not for absent objects (M_{diff_absent} = 0.153 – 0.155 =- 0.002, t < 1). These results demonstrate that both the levels of object presence and goal-relevance as manipulated in the current study resulted in different degrees of concept availability.

Discussion

We examined the quality of mental representations that are activated during language processing. Previous studies on perceptual representations manipulated visual object features to explicitly match or mismatch described object features. We did not. We manipulated the perceptual quality of the response options to the referent object and provided no matching decision alternative (both response options had a different transparency level than the original picture; thus, all answers were in principle incorrect). This manipulation revealed an implicit tendency of participants to select a more transparent (less vivid) picture after reading about a corresponding absent object. This implies that the entities that are not present in a described situation have a decreased perceivability in the referent representation. Studies that used probe words to assess the accessibility of visually absent, occluded, or distant objects found similar results in terms of availability of a corresponding concept (Glenberg, Meyer, & Lindem, 1987; Horton & Rapp, 2003; MacDonald & Just, 1989). Our current results extend these findings by specifying that a specific visual aspect of the concept representation is altered, namely its vividness.

To our knowledge, four previous studies have assessed perceptual quality (Hirschfeld & Zwitserlood, 2011; Holmes & Wolff, 2011; Risen & Critcher, 2011; Yaxley

& Zwaan, 2007), only one of which is directly relevant to our research question (Yaxley & Zwaan, 2007). Here, participants read sentences that manipulated the perceivability of an object (e.g., about a skier seeing a moose through fogged versus clean goggles), which facilitated response times to subsequently presented pictures of the object with a matching quality (low versus high visual resolution). Note that our current manipulation was more subtle. Our sentences did not mention perceived visual quality, they were embedded in stories in a task that did not require picture-to-sentence matching, and the pictures we presented in the experimental trials were not explicitly congruent or incongruent with the sentence in which the object was mentioned. Therefore, these findings are less susceptible to potential task-based strategies that participants might adopt to successfully perform the task[19].

Another difference is that we studied whether linguistic manipulations can qualitatively alter an experiential trace that was laid down within the experiment. All participants encoded the same pictures with identical (50%) transparency, thereby guaranteeing a baseline visual representation that was identical across conditions. Any difference in recollection of the encoded pictures should therefore be due to a difference in activation of this baseline trace. In other words, this paradigm taps into re-activation of a specific perceptual trace, rather than into a perceptual trace that might have been created ad hoc during language processing. For example, people may not have experience with seeing a moose through foggy goggles per se (Yaxley & Zwaan, 2007), but their (different) prior experience with the meaning of the words 'foggy' and 'moose' can still enable them to infer the approximate meaning of the sentence and envision a hazy moose. Therefore, it may be possible that people created a specific instant of a moose (of high or low visual quality), rather than recruiting a visual representation of a hazy moose from memory. For this reason, the current approach provides straightforward evidence that the *strengthening* of an experiential trace resulted in a *better quality* of the representation.

Whether a described object was relevant to the protagonist's goal did not affect the recalled transparency of encoded pictures. To ensure that this was not due to the stimuli we created, we performed a control experiment (Experiment 3) that demonstrated increased accessibility to relevant compared to irrelevant concepts, which confirms that our stimuli differentiated goal-relevance. A significant interaction revealed that the effect of relevance only occurred for stories in which the object was present. We hypothesized this pattern of concept availability based on the negation literature (hypothesis 1b) (Ferguson, Sanford, & Leuthold, 2008; Giora, 2006; Hasson & Glucksberg,

---

19 Any processing strategy in which participants did relate pictures to stories would have only been useful to distinguish experimental from filler stories, because this was the salient (but non-critical) manipulation.

2006; Kaup, Lüdtke, & Zwaan, 2006). Overall, this pattern suggests that the concept availability of absent objects was reduced to floor level, resulting in an effect of goal-relevance only for present objects.

Even though our control experiment and previous research (Albrecht & Myers, 1995; Dopkins, Klin, & Myers, 1993; Lutz & Radvansky, 1997; Magliano, Taylor, & Kim, 2005; Suh & Trabasso, 1993; Zwaan & Radvansky, 1998) convincingly showed that goal-relevance improves the accessibility of concept information, Experiments 1 and 2 did not reveal any effect of goal-relevance. This suggests that the current paradigm actually tapped into visual properties rather than overall concept availability or mere lexical accessibility, which supports hypothesis 2. It furthermore demonstrates that specific linguistic descriptions affect the mental representations we store in memory by differential reactivation of a perceptual trace. Based on these data alone, we cannot distinguish whether reading about absent objects did not recruit the stored perceptual information, or whether it did but the activation of the perceptual trace decayed more quickly. However, evidence from the negation literature suggests that early processing of negation/absence does not yet affect concept availability, only later processing does (Ferguson, Sanford, & Leuthold, 2008; Giora, 2006; Hasson & Glucksberg, 2006; Kaup, Lüdtke, & Zwaan, 2006). This suggests that the experiential trace of the referent concept might have been activated, but that it decayed (or was suppressed) shortly after. However, further research is needed to draw such conclusions within this framework.

People speak in terms of memories fading away, having a clear recollection of something, erasing images from one's mind, or having a situation fresh in memory. Our findings suggest that there is an actual physical component of mental representations that underlies these metaphors. In this study, we were able to isolate a qualitative visual component of concepts. This manifested itself in the recalled vividness of a stored representation. Even though objects were always explicitly mentioned in the text, the vividness of their corresponding representation in memory differed as a function of their described presence. These findings are in line with a grounded perspective on language processing, because they suggest that reading situation descriptions differentially re-activates experiential traces that were laid down previously.

The current approach provides new possibilities for future research. Given the wealth of demonstrations of perceptual involvement during language processing, we argue that at this point in time it is necessary to specify which aspects of concepts contribute to which aspects of language comprehension. This way, we will be able to formulate the strengths and weaknesses of the grounded (experiential) view on language processing and take new steps in understanding how people comprehend language and represent the external world.

Acknowledgements

We thank Lera Boroditsky for facilitating the experimental procedure and for her helpful comments on the research design. Furthermore, we thank Jan Engelen, Nicole Goossens, Jim Maarseveen, Jacqueline de Nooijer, Lysanne Post, and Karen Schuil from the Language and Cognition Lab at Erasmus University Rotterdam for their help with the materials and design. Also, we thank Jane L. Risen for additional information about Experiment 6 of the Risen and Critcher paper (2011). Finally, we thank four anonymous reviewers for their valuable comments on an earlier version of the manuscript.

Table S1. Online Appendix A.

Experimental stories. Object presence is indicated by the articles ‘a’ or ‘no’ in sentence 4 of each story. Goal-relevance is presented in different columns: The left column contains stories in which the object is relevant; the right column contains stories in which the object is irrelevant.

Table S1. Online Appendix A.

Relevant	Irrelevant
Charlotte was stranded on a deserted island.	Charlotte was walking down the beach.
After a few days there was almost no food left.	It was early in the morning.
She desperately scanned the horizon	Both the beach and the ocean were remarkably quiet.
Charlotte saw a/no *boat* sailing by.	Charlotte saw a/no *boat* sailing by during her walk.
The sun was beating down.	Just as Charlotte wanted to go home, it started to rain..
Linda went out for lunch with her colleagues.	Linda went out for lunch with her colleagues.
They found a nice restaurant with a beautiful view.	They found a nice restaurant with a beautiful view.
When they sat down, Linda found herself very thirsty.	The waitress came to take their orders.
She saw that there was a/no *bottle of water* on the table.	Linda saw that there was a/no *bottle of water* on the table.
After Linda and her colleagues finished their lunch they went back to the office.	After they finished their lunch they went back to the office.

Relevant	Irrelevant
For his new job, Jimmy had to move to San Francisco.	To pay his rent, Jimmy worked night shifts in a factory.
He hired a moving company to move all his stuff.	After three hours of strenuous work, Jimmy had a break.
Half an hour after arrival of the company, Jimmy checked whether they started already.	Another worker took over for the time being.
When he looked into the truck, there was a/no *box*.	When he came back, there was a/no *box* on the conveyor belt.
An hour later he went to his landlord to hand in the keys.	Jimmy realized he had come back from his break too late.
Today was Mel's first day on the job as a secretary.	Today, Mel started her new job as a secretary.
She was in the copy room to print some forms.	She started out by checking her email and making tea for her boss.
Suddenly, she felt dizzy.	Then she walked to the copying room to print some forms.
She noticed that there was a/no *chair* in the copy room.	She noticed that there was a/no *chair* in the copying room.
Later that day, she called the doctor for an appointment.	Then she realized that she forgot to bring her boss his tea and hurried back to the kitchen.

Fiona wanted to surprise her mom for Mother's Day.	Fiona wanted to surprise her mom for Mother's Day.
She entered the house using the key she kept and started cooking dinner.	She entered the house using the key she kept and started cooking a nice dinner.
She thought it would be nice to start with a hot beverage.	When the main course was almost done, she waited and looked around.
She saw a/no *coffee machine* in the kitchen.	She saw a/no *coffee machine* in the kitchen.
Then the timer sounded and she took the rice off the stove.	Then the timer sounded and she took the rice off the stove.
Keith was working on his term paper in the computer room.	Keith was working on his term paper in the computer room.
After some time he craved a cigarette.	After some time he craved a cigarette.
On the way out, he walked to the vending machine in the hallway.	On the way out, he checked the vending machine in the corridor.
He found a/no *coin* in his wallet.	He found a/no *coin* left behind in the change slot.
Then he heard a familiar voice behind him.	Then he heard a familiar voice from behind him.

Relevant	**Irrelevant**
Peter was a plumber.	Peter was driving around in his delivery van.
He stopped in front of a big house for his first job of the day.	He had to deliver some packages before the end of his shift.
The sidewalk was very muddy.	He stopped in front of a big house.
As he walked up to the door, he saw a/no *doormat*.	As he walked up to the door, he saw a/no *doormat*.
He rang the doorbell.	Then suddenly, a dog started barking and Peter ran back to his van.
Jake and Sally were hiking in the hills.	Jake and Sally were hiking in the hills.
They enjoyed the sunny weather as they came across a meadow with grazing cows.	They kept a swift pace and enjoyed the sunny weather.
Sally was afraid the cows would come too close.	After several hours, they came across a quaint little pasture.
There was a/no *fence* bordering the meadow.	There was a/no *fence* bordering it.
Just then, the clouds rolled in to cover the sun.	Just then, the clouds rolled in to cover the sun.

Josh and Mary had rented a vacation home.	Josh and Mary had rented a vacation home.
The house had a hot tub, wireless internet, and a fireplace.	It was situated beautifully at the top of a mountain.
It was cold at night.	The house had a hot tub, wireless internet, and a fireplace.
There was (no) *firewood* in the house.	There was (no) *firewood* in the house.
Josh and Mary opened up a bottle of red wine.	Josh and Mary enjoyed the hot tub.
Fred was waiting at the hairdresser's.	Fred was waiting at the hairdresser's.
He was the third person in line for a haircut.	He was the third person in line for a haircut.
Fred was bored and wanted something to read.	He looked around and observed the other customers.
There was a/no *magazine* on the table.	There was a/no *magazine* on the table.
A few minutes later, it was Fred's turn already.	A few minutes later, it was Fred's turn already.

Relevant	Irrelevant
Annie was on a long hiking tour.	Annie was on a long skating tour.
After a few hours, she started to feel unwell.	After one hour she stopped to take a rest.
Her throat was aching.	Her legs hurt a little.
She had a/no *peppermint* in her pocket.	She had a/no *peppermint* in her pocket.
Then her mobile phone went off.	After ten minutes she started skating again.
Bob was invited to a birthday party on Friday evening.	Bob arrived at a birthday party on Friday evening.
He went to the party straight from work.	It had been a very busy week at work.
He did not have time for dinner so he was hungry.	He was tired and wanted to sit down.
At the party there was a/no *piece of pie* left for him.	There was a/no *piece of pie* left for him.
After a couple of hours, Bob went home and fell asleep quickly.	After about an hour, Bob went home and fell asleep quickly.

Joan was on holiday in a small town in Egypt.

She went diving almost every day.

But she was also dying to see some of the ancient Egyptian architecture.

In the small town there was a/no *pyramid*.

Joan had decided on coming back next year.

Jenna planned on going to the beach, since the weather was lovely.

She packed her bag and drove to the coast.

Jenna liked collecting pretty things from the beach.

When her feet touched the water, she looked down where she saw a/no *sea shel*l.

The water wasn't as warm as she hoped and she walked back to her towel.

Joan was on holiday in a small town in Egypt.

She went diving almost every day.

In the evenings she went to her favorite restaurant.

In the small town there was a/no *pyramid*.

Joan had decided on coming back next year.

Jenna planned on going to the beach, since the weather was lovely.

She packed her bag and drove to the coast.

Once she got there, she decided to go for a swim in the ocean straight away.

When her feet touched the water, she looked down where she saw a/no *sea shell*.

The water wasn't as warm as she hoped and she walked back to her towel.

Relevant	Irrelevant
Milly took her little nephew to the zoo.	Milly took her little nephew to the zoo.
They ended up at a little playground.	He was scared of the lions, but loved the monkeys.
After a while Milly's nephew was tired of the swings and wanted to play somewhere else.	At the end of the day they ended up at a little playground.
Milly saw that there was a/no *slide* there.	Milly saw that there was a/no *slide* there.
After playing for a while Milly brought her nephew back home.	After playing for a while Milly brought her nephew back home.
Sam was about to go out for a walk.	Sam just woke up and was about to have breakfast.
His roommate asked Sam to post his letter.	He was hungry, so he decided to make some toast.
Sam grabbed the envelope to make sure the postal charges were covered.	Then he noticed an envelope on the table with his name on it.
There was a/no *stamp* in the upper right corner.	There was a/no *stamp* in the upper right corner.
Then he realized his friends would come over today.	Before he could open the envelope, a weird smell made him realize that his toast was burned.

Will went to help his brother paint his new house.

It was pretty warm that day.

At lunch they wanted to sit outside in the shadow.

There was a/no *tree* in the backyard.

Will offered to begin with the living room.

Jennifer was on a mountain bike trip with her friends.

After an hour, everyone had gotten thirsty.

They stopped to take a break at a picnic area.

Jennifer saw a/no *water fountain* there.

After fifteen minutes, the group climbed back on their bikes.

Will went to help his brother paint his new house.

It was old and needed a lot of work, but it was a cozy place.

When they had enough of it, they sat down and relaxed in the back yard.

There was a/no *tree* there.

Will offered to come back the next day to finish the work.

Jennifer was on a mountain bike trip with her friends.

The terrain was hilly and the trails were difficult.

After an hour, they stopped to take a break at a picnic area.

Jennifer saw a/no *water fountain* there.

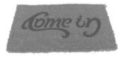
After fifteen minutes, the group climbed back on their bikes.

Relevant	**Irrelevant**
Vanessa walked down the supermarket aisle.	Vanessa was doing some shopping.
She wanted to make a dessert with fruit.	She wanted to make a dessert for her friends.
The recipe required exactly 200 milligrams of black berries and cherries.	She walked to the storing shelf with fresh fruit.
She looked around and saw there was a/no *weighing scale*.	There was a/no *weighing scale*.
Ten minutes later she walked out of the supermarket.	Fifteen minutes later she walked out of the supermarket.
Bill landed at the airport late at night.	Bill landed at the airport late at night.
An hour later, he checked into his hotel room.	An hour later, he checked into his hotel room.
Bill felt like having a drink before going to sleep.	Bill decided to relax a little before going to sleep.
He saw that there was a/no *whiskey bottle* in the minibar.	He saw that there was a/no *whiskey bottle* in the minibar.
Then he remembered he promised to call his wife upon arrival.	Feeling quite exhausted after all, he went to the bathroom to brush his teeth.

Table S2. Online Appendix B.

Experimental pictures with different transparency levels. From left to right: 45%; 50%; 60%.

Object	45%	50%	60%
boat			
bottle of water			
box			
chair			
coffee machine			
coin			

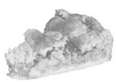
doormat   

fence   

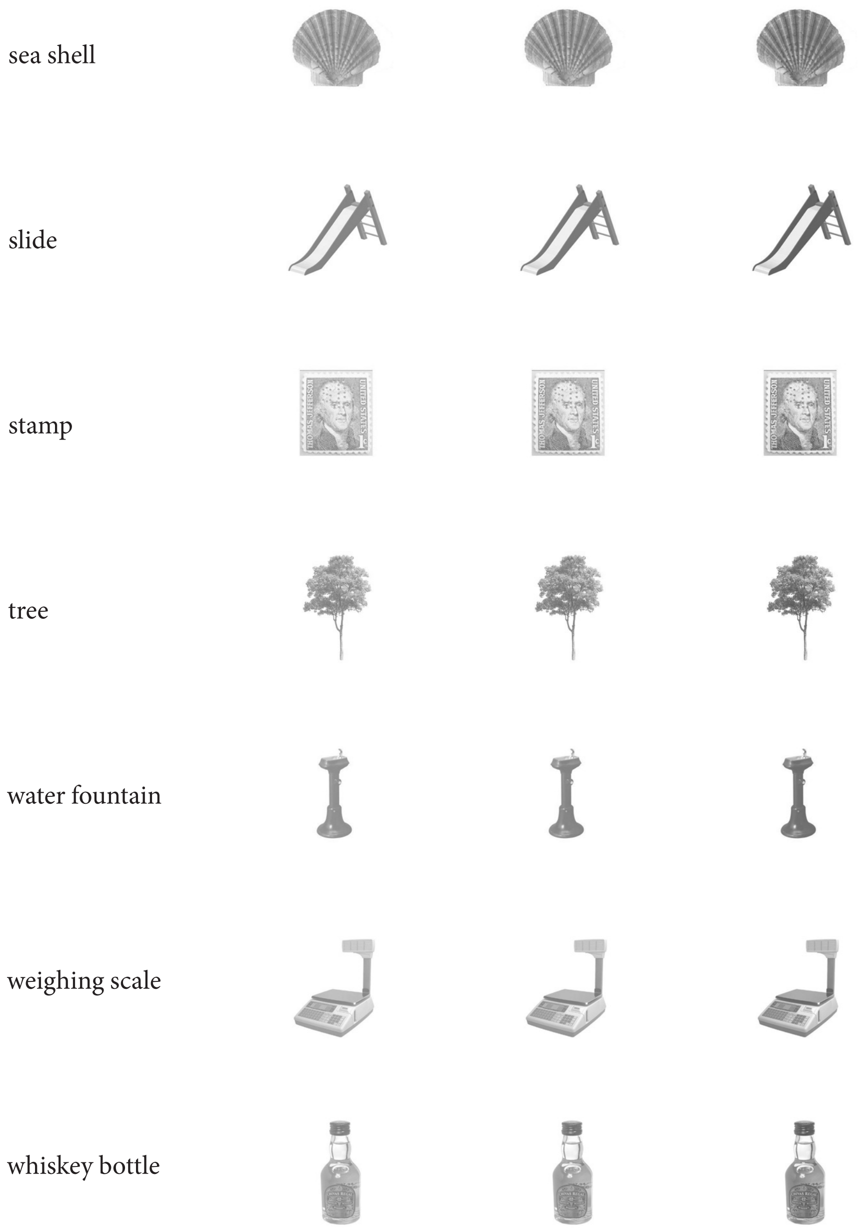
firewood   

magazine 

peppermint   

piece of pie   

pyramid   


sea shell
slide
stamp
tree
water fountain
weighing scale
whiskey bottle

Chapter 5

Seeing the apple in the fruit: Implied linguistic specificity affects the detail of visual representations

Abstract

Early work in cognitive psychology shows that people encode general (category) information in terms of specific exemplars that are implied by the linguistic context. We combine this insight with a grounded cognition perspective that lays down a role for perceptual involvement in language comprehension. We examine whether linguistic specificity interacts with specificity in the visual domain, and hypothesize that instantiating exemplars for general terms should be accompanied by more detailed visual information than conceptualizing the general terms themselves. Experiment 1 replicates early work by showing that specific lexical cues such as 'apple' (rather than general lexical cues such as 'fruit') result in better retrieval for sentences biasing towards that instantiation ('The fruit forms an ingredient for cider') than for sentences that avoid this bias ('The fruit forms an ingredient for exotic salads'). Experiment 2 extends these findings to the visual domain, by showing that detailed line drawings (rather than schematic outlines) provide better cues for biasing than broad sentence contexts. This suggests that different description levels are associated with different levels of conceptual visual detail. Experiment 3 confirms this interpretation of the findings, by showing that online language processing evokes different degrees of visual specificity. These results support findings from both cognitive linguistics and grounded cognition, and provide novel insights about the specificity with which visual information of a concept is stored and retrieved during language processing.

This chapter is submitted as: Vandeberg, L., & Zwaan, R.A. (subm). Seeing the apple in the fruit: Implied linguistic specificity affects the detail of visual representations.

Introduction

Reading that 'Paul McCartney played his instrument' will likely make you think of a different instrument than reading that a random 'musician played his instrument', or 'Ringo Starr played his instrument', or even 'Jimi Hendrix played his instrument'. Similarly, the word 'playing' should be interpreted differently when being in the context of a bass guitar, a piano, a casino, or a playground. This shows that rather than having one meaning, words can have a whole *family* of meanings (Wittgenstein, 1953, in Mervis & Rosch, 1981). Early research on semantic flexibility has shown that words have different nuances of meaning, depending on the context in which they appear (Anderson & McGaw, 1973; Anderson & Ortony, 1975; Anderson et al., 1976; Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974; Halff, Ortony, & Anderson, 1976; Swinney, 1979; Swinney & Hakes, 1976; Tabossi & Johnson-Laird, 1980). Particularly, Anderson and colleagues (1976) have found that the level of specificity suggested by a sentence context affects the degree of specificity with which the concept is stored in memory.

In their study on specificity, Anderson and colleagues (1976) created sentence pairs with identical subject nouns (e.g., 'woman') in different contexts. Target sentences biased towards a specific instantiation of the general subject noun (e.g., 'The woman was outstanding in the theater', suggesting an actress) whereas control sentences did not (e.g., 'The woman worked near the theater'). Participants read the sentences and were later given a cued-recall test, with two cues for each sentence. One cue provided the particular term that was suggested by the biasing context ('actress'), the other cue was the presented general term ('woman'). Participants were instructed to recall the last word of the previously presented sentence. The assumption here was that recall depends on what people have encoded and stored (referred to as *encoding-specificity*) (Reder, Anderson, & Bjork, 1974; Tulving & Thomson, 1973). Results showed that the cue naming the suggested instantiation ('actress') was substantially better at evoking the target sentence than the general cue ('woman'), even though it had not been explicitly presented in the to-be-remembered sentence. This was taken as evidence that general terms are instantiated, meaning that they are encoded on the basis of the exemplars that are implied by the context.

Other evidence for the activation of exemplars when processing general (category) terms is found in studies on anaphoric reference (e.g., Corbett & Chang, 1983; Garrod & Sanford, 1977; Gernsbacher, 1989). For example, Garrod and Sanford (1977) assessed reading times when participants read the last sentence of passages in which categories were mentioned after instances (e.g., 'A *bus* came roaring round the corner. The *vehicle* narrowly missed a pedestrian') or before instances ('A *vehicle* came roaring round the corner. The *bus* narrowly missed a pedestrian'). They found shorter reading times for the category-last passages than for the category-first passages, showing that it was easier to integrate 'vehicle' after reading about a bus than vice versa. The authors take these findings to suggest that terms are represented at the most specific level possible. As a result, reading the last sentence of the category-first condition would require a change in specification (*bus* contains novel specific information that needs to be to be incorporated), but not in the category-last condition (*vehicle* can be represented at the *bus* level of specificity which was already established in the first sentence).

Thus, evidence shows that the construal of general terms consists of activating specific concepts, at least whenever the context biases towards an instantiation. All this research has been performed within the realm of language, using lexical stimuli, lexical cues, and/or 'lexical' measures such as reading times or word retrieval. However, concepts do not only consist of lexical information, but also have perceptual and functional attributes and are shaped by previous experience and world knowledge (Murphy & Medin, 1985). Although early theorists have speculated about a role for functional and perceptual information of concepts (Jackendoff, 1987; Lakoff, 1990; Langacker, 1986; Talmy, 1988), it was not until the highly influential paper of Barsalou (1999) that researchers have started to look for direct evidence for these attributes in language research. Barsalou claimed that in order to derive meaning from the abstract symbols that language consists of, language should be grounded in bodily experience. This harmonizes with the view that in order to comprehend language, one needs to represent the described situation rather than the text itself (Bower & Morrow, 1990; Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). A large body of evidence from the past two decades supports this claim of grounded cognition by demonstrating that visual and motoric information is involved in language processing. One of the earliest studies on the interaction between visual and linguistic processes demonstrated that readers were faster to identify pictures in which the visual features of an object were congruent rather than incongruent with previously described events (Stanfield & Zwaan, 2001). A variety of studies have since found perceptual recruitment in language using different paradigms, stimuli, and dependent variables (Fincher-Kiefer, 2001; Horton & Rapp, 2003; Lupyan & Thompson-Schill, 2012; Richardson, Spivey, Barsalou, & McRae, 2003). Visual features such as shape, orientation, color, size, and (implied) motion direction are represented when processing language about objects or actions (Borghi & Riggio, 2009; Connell, 2007; Dils & Boroditsky, 2010; Hirschfeld & Zwitserlood, 2011; Holt & Beilock, 2006; Kaschak et al., 2005; Lincoln, Long, & Baynes, 2007; Stanfield & Zwaan, 2001; Zwaan, Madden, Yaxley, & Aveyard, 2004; Zwaan, Stanfield, & Yaxley, 2002).

With this evidence for grounded cognition in mind, the fact that a biasing context evokes the instantiation of a suggested exemplar can be expected to leave its traces in the visual domain. In this case, the exemplar should be associated with greater visual specificity than the category (which leaves the options for instantiations open). This line of reasoning is broadly consistent with that by Langacker, who -in his commentary on the 1999 Barsalou paper- suggested that the degree of specificity in description levels may be analogous to that of their referent visual experiences:

> *'(…) linguistic expressions can characterize a situation at any desired level of precision and detail ('granularity'), as seen in lexical hierarchies such as thing – creature - reptile - snake - rattlesnake - sidewinder. This progression from highly schematic to increasingly more specific expressions seems quite analogous to the visual experience of seeing something with progressively greater acuity while walking up to it from a distance.' (p. 625)*

To our knowledge, no empirical research to date has examined this notion. Given the large amount of research on the interactions between linguistic and visual information, it is remarkable that our knowledge about the visual component of concepts is still very much limited to contrasting visual features of described objects and actions, such as horizontal versus vertical orientation or movement (but also see the few studies that have recently started to assess the acuity of visual information that accompanies language processing (Hirschfeld & Zwitserlood, 2011; Holmes & Wolff, 2011; Vandeberg, Eerland, & Zwaan, 2012; Yaxley & Zwaan, 2007). By combining the insights from grounded cognition and cognitive linguistics, we can start examining the generalizability and limitations of the findings from both fields. With this purpose, we aimed to examine whether specificity in the linguistic domain interacts with specificity in the visual domain. We hypothesized that the instantiation of a general term should be associated with greater visual specificity than conceptualizing the general term itself.

In Experiment 1, we adopted the encoding-retrieval paradigm with word cues from Anderson et al. (1976). We created novel sentence stimuli with novel specific and general word cues that could be transformed into detailed and schematic line drawings for Experiment 2. Experiment 2 was similar to Experiment 1, except that we used specific and general line drawings as cues rather than using words. By testing the encoding-retrieval paradigm across modalities we tested how visual (line drawing) cues affect sentence retrieval. In this manner, we explored whether visual specificity interacts with linguistic specificity. Finally, Experiment 3 employed a sentence-picture verification paradigm, with which we assessed whether an interaction between linguistic and visual specificity emerges during online language processing, i.e., whether it occurs during encoding rather than during retrieval. Such a finding would suggest that readers represent instantiations with greater visual detail than categories during language processing, in a way that is congruent with Langacker's notion.

In creating the novel stimuli for our experiments we were unable to keep the last word in each sentence of a pair identical and were therefore in need of a different dependent measure than that of Anderson et al. (1976). Experiment 1 served to (a) test whether our stimuli and dependent measure were able to tap into different levels of specificity in participants' responses, (b) conceptually replicate Anderson's findings within the linguistic domain, and (c) obtain a baseline retrieval pattern as a result of word cues with which we could compare that of line drawing cues from Experiment 2.

Experiment 1

Method

Ethics Statement

All participants were recruited online and voluntarily subscribed for participation. Written consent was not obtained because the norming studies and experiments were noninvasive. This is in accordance with departmental practice approved by the Ethics Committee of Psychology (ECP) at the Erasmus University Rotterdam, the Netherlands.

Norming study

A norming study was designed to select the stimuli. We created 31 sentence pairs. Each pair consisted of two sentences with the same subject noun (e.g., 'fruit'), one in a biasing context ('The fruit forms an ingredient for cider') and one in a broad context ('The fruit forms an ingredient for exotic salads'). The biasing context directed the interpretation of the category subject noun towards a specific instantiation ('apple'), whereas broad contexts avoided such a bias. Two lists were created that each contained one sentence of each pair. In each list, half of the sentences had a biasing context and the other half had a broad context. Eighty-four participants were recruited online through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 30 (ranging from 18-64, SD = 10) and contained 52 females (62%). They were paid $0.50 for their participation, which took approximately 8 minutes. They were asked whether a specific example of the category subject noun came to mind. They were instructed to enter an example if one came to mind and to enter the subject noun if no example came to mind. Results showed that participants rarely entered the category nouns (M = 5%, SD = 5%). We selected 20 sentence pairs of which the biasing context elicited one single example frequently (M = 77%, SD = 12%) and the broad context elicited a variety of responses that never or rarely included the target example (M= 4%, SD = 5%).

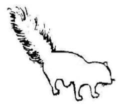
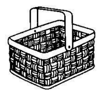
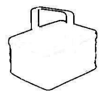
Participants

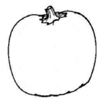
We recruited 199 participants online through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 33 (ranging from 15-64, SD = 11) and contained 125 females (63%). All participants were residents of the USA and native speakers of English. They were compensated with $1.50 for their participation, which required approximately 25 minutes.

Stimuli

Each of the selected 20 items consisted of one sentence with a biasing context ('The fruit forms an ingredient for cider'), one sentence with a broad context ('The fruit forms an ingredient for exotic salads'), one specific word cue (apple), and one general word cue (fruit). See Table 1 for some sample items. The sentences were counterbalanced across two lists, with one list consisting of 10 biasing and 10 broad contexts, and the other list consisting of the other 10 biasing and 10 broad contexts. Analogously, word cues were divided across two blocks, in which each block contained 10 specific and 10 general word cues. The context lists were counterbalanced between subjects and the cue blocks were counterbalanced within subjects, resulting in 4 counterbalancing lists.

Table 1. Sample stimuli

Cell	Sentence	Exp. 1 Word Cue	Exp. 2 Picture Cue
Bi_S	The small animal is characterized by its repulsive smell.	Skunk	
Br_S	The small animal is characterized by its diet of nuts and insects.		
Bi_G	The small animal is characterized by its repulsive smell.	Small animal	
Br_G	The small animal is characterized by its diet of nuts and insects.		
Bi_S	The container that held the picnic utensils was left in the car.	Basket	
Br_S	The container that held the equipment was left in the car.		
Bi_G	The container that held the picnic utensils was left in the car.	Container	
Br_G	The container that held the equipment was left in the car.		

Bi_S	The vegetable for halloween decorations was bought on the market.	Pumpkin	
Br_S	The vegetable for salads and stews was bought on the market.		
Bi_G	The vegetable for halloween decorations was bought on the market.	Vegetable	
Br_G	The vegetable for salads and stews was bought on the market.		

Note. The cell column contains information about the conditions, with Bi_S representing the condition with the Biasing sentence and Specific cue; Br_S being Broad sentence and Specific cue; Bi_G being Biasing sentence and General cue; Br_G being Broad sentence and General cue.

Procedure

The experiment was programmed and presented in the online Qualtrics survey research suite (http://www.qualtrics.com). Participants performed a task that consisted of three parts and that was very similar to that of Anderson et al. (1976). In the first part, participants were randomly presented with the twenty sentences at an automatic pace of 7 seconds per sentence. Participants were instructed to carefully read and remember the sentences because details would be requested in a later part of the experiment.

Next, participants performed a short and unrelated filler task in which they were instructed to complete 10 logical sequences. The sequences consisted of 3 grids in which small circles formed a pattern. The participants were instructed to select out of four grids the one that would logically follow the first 3 grids. If they had not responded within 30 seconds, the next sequence would be presented. This way, the filler task would take a maximum of about 6 minutes. The purpose of this filler task was to minimize recall from short-term non-semantic memory, cf. Anderson et al. (1976).

Finally, participants performed a cued-recall test. They received two cues for each presented sentence; one general word cue that consisted of the category name (e.g., fruit) and one specific word cue that consisted of the example (e.g., apple). The cues were counterbalanced across blocks, so that the first block contained one cue of an item and the second block contained the other cue of that item. The cues were presented individually and were randomized within a block. Participants were instructed to reproduce the sentence of which the cue reminded them and were warned that this could result in producing certain sentences more than once. They were encouraged to be as accurate as possible, even though they might not recall the exact sentences. Once they had entered the sentence in the space on the screen, they had to press the >> button on the bottom of the screen to continue with the next trial. After the task, participants were queried about their notion of the purpose of the experiment and filled out a brief demographic questionnaire.

Analysis

We assessed recall of the sentences by means of Latent Semantic Analysis (LSA: http://lsa.colorado.edu). LSA is a technique that can be used for statistically computing the semantic similarity between words or texts (Landauer & Dumais, 1997). It calculates the contextual distance between words or texts that are represented in a multidimensional space that is based on a large text corpus. LSA has been shown to mimic human language structures in a variety of ways (Landauer, Foltz, & Laham, 1998). By means of LSA, we calculated the semantic overlap between the presented sentences (e.g., 'The fruit forms an ingredient for exotic salads') and those that were produced by the participants (e.g., 'The fruit is often used in exotic salads'), in which higher cosine values represent a higher degree of semantic overlap. We used this measure based on the assumption that a greater semantic relatedness between the presented and produced sentences should indicate that participants recalled the (gist of the) sentence to a greater extent.

Results

The computed LSA cosines ranged from -0.12 to 1, with the lowest cosines indicating that the produced sentence showed no semantic overlap with the presented sentence whatsoever (often as a result of belonging to a different item) and the highest cosine of 1 indicating a perfect match between the presented and the produced sentence.[20] Participants who had only responded with single words rather than sentences (N= 13) and participants with extremely low average LSA cosines across conditions (<.20, N= 9, of which 5 also belonged to the single word group) were excluded from the analysis, because these participants did not follow the instructions and/or showed extremely low recall. To obtain equal sample sizes per counterbalancing list, 10 additional participants (being 5, 3, and 2 participants in one of each lists) who were recruited last in their list were excluded from the data. Exclusion of all 27 participants did not alter the pattern of results or any of the significance values. That is, all of the reported significant effects were already significant and in the same direction before this data removal.

We averaged the LSA cosines across items in a condition for each participant, see Figure 1. A 2 (context) x 2 (word cue) repeated-measures ANOVA revealed a significant main effect of context type, which demonstrated higher LSA cosines for the biasing (M = .49) than the broad (M = .41) contexts, $F(1,171) = 49.79$, $p < .001$, $\eta^2 = .23$. It further showed a significant main effect of word cue type, with higher LSA cosines for general (M = .51) than for specific (M = .40) cues, $F(1,171) = 84.06$, $p < .001$, $\eta^2 = .33$. Importantly, a significant interaction ($F(1,171) = 171.20$, $p < .001$, $\eta^2 = .50$) showed that the context effect only occurred for the specific word cues. Post-hoc two-tailed paired-samples *t*-tests showed that the specific words were better cues for the sentences with biasing compared to broad contexts (t (171) = 12.70, $p < .001$, $r = .47$). The general terms showed no such difference (t (171) = 1.95, $p = .053$, $r = -.07$). Though this latter effect is marginally significant, we will not interpret it as meaningful. Given the extremely small effect size, the *p*-value almost reaching significance is likely the result of the high power resulting from our large sample.

20 Given that a cosine of 0 reflects no semantic overlap whatsoever, it is difficult to grasp the meaning of negative cosines. For this reason, we have checked whether negative cosines may have affected the pattern of results in some way. We performed an analyses on the data in which negative cosines were transformed into 0 cosines. Due to the power, the pattern of results remained identical to that described in this section. Therefore, all described analyses are based upon the original LSA cosines including the negative values.

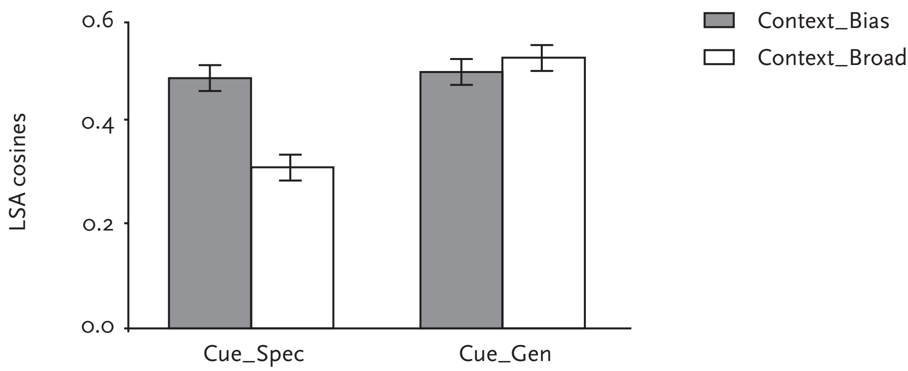


Figure 1. LSA similarity scores between presented and produced sentences with word cues. Error bars represent the 95% confidence intervals for each condition.

The main finding lies in the interaction, showing that specific words serve as a better cue for biasing than for broad sentences. In line with Anderson et al. (1976), we argue this to be the result of people instantiating the implied exemplar in biasing rather than broad contexts. However, an alternative explanation to this finding could be that participants had equally low recall for both types of sentence context when using a specific cue, but that they simply made a guess based on the specific cue which would likely result in a higher overlap with the biasing than the broad context. For example, in case of the 'apple' cue, if people did not recall the sentence and guessed, they might be more likely to include (associations with) 'cider' in their sentence than (associations with) 'salad'. If this reasoning is correct, the produced sentences in the broad context should provide higher cosines with the biasing (not presented) rather than the broad (presented) context. In order to check whether this was the case, we calculated the semantic overlap between the produced sentences in the broad context - specific cue condition and the biasing sentences. Two-tailed paired-samples *t*-test showed that the overlap with biasing sentences (M = 0.29) was even worse than that with broad sentences (M = .31), though with a very small effect size (t (171) = 3.26, p = .001, r = 0.06). This result rules out the proposed alternative explanation that participants simply made a guess based on the specific cue rather than that they recalled the broad context.

Discussion

The results strongly resemble those of Anderson et al. (1976), whose main finding was higher recall for biasing than broad contexts when using a specific cue. Note that the specific cue was never presented in a sentence. Therefore, any effect of the specific cue on retrieval indicates that the cued exemplar was instantiated upon encoding the sentence.

The difference between our results and those of Anderson and colleagues lies in the general cues. Whereas they found that recall with the general retrieval cues was relatively low (similar to the lowest recall condition of the specific cues), it is relatively high in our study (similar to the highest recall condition of the specific cues). This is likely the result of our dependent measure. Our LSA recall measure is calculated on the entire produced sentence that necessarily includes the general term, whereas Anderson's recall rate was based on the final word of the sentence which did not include the general term. Thus, mentioning the general cue in a recalled sentence increases the overlap between the presented and produced sentences relative to calculating a recall rate to an unrelated word. Note that this is not problematic for our findings, given that the relevant test lies in the interaction and not in the main effect of cue.

With different stimuli and a different dependent measure, but with the same experimental paradigm and pattern of results, we have conceptually replicated the findings by Anderson et al. (1976). Both studies showed that specific examples provide a better retrieval cue for sentences biasing towards that example than for sentences that did not. Given the large amount of evidence that people's conceptual system contains more than solely symbolic lexical information, we now aimed to assess whether these findings generalize to the visual domain.

Experiment 2

Based on the theoretical (Barsalou, 1999; Jackendoff, 1987; Lakoff, 1990; Langacker, 1986; Talmy, 1988) and empirical (Fincher-Kiefer, 2001; Horton & Rapp, 2003; Lupyan & Thompson-Schill, 2012; Richardson et al., 2003; Stanfield & Zwaan, 2001) literature about visual involvement in constructing mental representations in language, we hypothesized that linguistic and visual specificity should interact. In this case, the retrieval of biasing contexts would benefit from specific (versus schematic) visual cues more than the retrieval of broad contexts. Such an interaction pattern would be comparable to that of Experiment 1, showing a visual specificity advantage for biasing rather than broad linguistic contexts. Alternatively, a lack of significant interaction would demonstrate that there is no interplay between specificity in the linguistic and the visual domain.

Method

Norming study

We performed a norming study on the line drawings that would replace the word cues, to confirm whether the schematic line drawings were less directive with respect to the implied exemplar than the detailed drawings. Two lists of 20 line drawings were created, each of which contained one picture of a pair. In each list, half of the pictures were detailed (specific) line drawings depicting an instantiation, and half were schematic (general) line drawings from which details were erased that were characteristic for the instantiation. Eighty participants were recruited online through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 34 (ranging from 18-65, SD = 12) and contained 35 females (44%). They were paid $0.40 for their participation, which took approximately 7 minutes. They were asked what the line drawing depicted and were instructed to enter the examples that came to mind (with a maximum of eight examples). Results showed that general line drawings elicited more examples (e.g., apple and/or peach and/or plum) than the specific line drawings (e.g., apple), t (79) = $p < .001$ (M = 1.89 and 1.29, respectively).

Participants

We recruited 198 new participants that had not participated in the norming studies or Experiment 1 through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 33 (ranging from 18-65, SD = 11) and contained 133 females (68%). All participants were residents of the USA and native speakers of English. They were compensated with $1.50 for their participation, which required approximately 25 minutes.

Stimuli

The stimuli and lists were identical to those of Experiment 1, with exception of the cues. Instead of presenting word cues, we presented participants with picture cues. These picture cues were detailed line drawings that served as specific cues, from which several lines were erased to create general cues, see Table 1.

Procedure

The procedure was identical to Experiment 1, the only exception being that the word cues were replaced by picture cues. In the cued-recall test, participants were instructed to reproduce the sentence of which the line drawing reminded them.

Results

The computed LSA cosines ranged from -0.11 to 1.[20] Participants who had only responded with single words rather than sentences (N= 10) and participants with extremely low average LSA cosines (<.20, N= 17, of which 4 also belonged to the single word group) were excluded from the analysis, because these participants did not follow the instructions and/or showed extremely low recall. To obtain equal sample sizes per counterbalancing list, 7 additional participants were excluded from the data (i.e., the 3, 3, and 1 last participant(s) were removed from one of the lists). Again, exclusion of all of the above participants did not affect the pattern of results or any of the significance values.

Again, we averaged the LSA cosines across items in a condition for each participant, see Figure 2. A 2 (context) X 2 (picture cue) repeated-measures ANOVA revealed a significant main effect of sentence type, which demonstrated higher LSA cosines for the biasing (M = .48) than the broad (M = .36) sentence contexts, $F(1,167) = 103.47$, $p < .001$, $\eta^2 = .38$. There also was a significant main effect of picture cue type, with higher LSA cosines for specific (M = .44) than for general (M = .40) cues, $F(1,167) = 100.79$, $p < .001$, $\eta^2 = .38$. Importantly, there was a significant context * picture cue interaction ($F(1,167) = 24.27$, $p < .001$, $\eta^2 = .13$), which showed that the context effect was significantly larger for specific than for general picture cues. Post-hoc two-tailed paired-samples *t*-tests showed that the specific pictures were better cues for the sentences with biasing compared to broad contexts ($t(167) = 12.14$, $p < .001$, $r = .43$). The general cues showed a difference in the same direction ($t(167) = 7.09$, $p < .001$, $r = .31$). These results show that the retrieval of biasing contexts benefitted from specific (versus schematic) visual cues more than the retrieval of broad contexts.

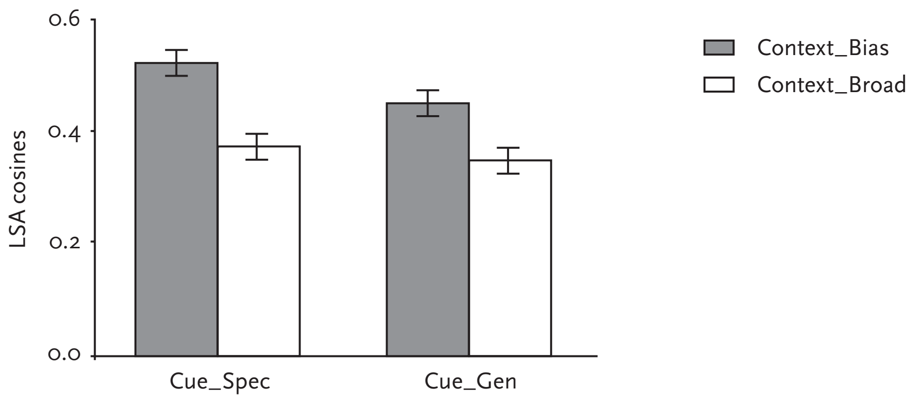


Figure 2. LSA similarity scores between presented and produced sentences with picture cues. Error bars represent the 95% confidence intervals for each condition.

Although the advantage of the specific cue mainly manifested for biasing contexts in both experiments, the pattern of results in Experiment 2 is different from that of Experiment 1. A 2 (context) x 2 (cue) x 2 (experiment) mixed design ANOVA revealed a significant three-way interaction ($F(1,338) = 65.22$, $p < .001$, $\eta^2 = .16$), which confirmed that the context*cue interaction patterns differed between the two experiments. This difference between experiments will be discussed in detail in the discussion section.

To check again whether the produced sentences in the broad context might provide higher LSA cosines to the biasing (not presented) rather than the broad (presented) context, we computed the semantic overlap between the produced sentences in the broad-context – specific cue condition and the biasing sentences. Two-tailed paired-samples *t*-test showed again that the overlap with biasing sentences ($M = 0.33$) was even worse than that with broad sentences ($M = 0.37$, $t(167) = 7.16$, $p < .001$, $r = .14$), which rules out the alternative explanation that participants simply made a guess based on the specific cue rather than recalled the broad context.

Discussion

The current results resemble those of Experiment 1, in the sense that they show a significant two way interaction between context specificity and cue specificity, with an advantage for the specific cues (e.g., an apple) when retrieving a biasing ('cider') compared to a broad ('salad') context. This supports our hypothesis by showing that specificity in the visual domain interacted with specificity in the linguistic domain. We discuss the implications of these findings in terms of the differences between Experiments 1 and 2.

The first difference between the two experiments is that the general picture cues (like the specific cues), showed a greater recall of biasing than broad contexts, which was not the case in Experiment 1. This context effect for general picture cues was unexpected, but can be explained by an overall advantage for biasing sentences. Given that this advantage for biasing sentences is relevant to all our findings, its implications will be thoroughly discussed in the general discussion.

The second difference between the two experiments lies in the main effect of cue type. Whereas the general word cues resulted in higher recall than the specific word cues in Experiment 1, the general picture cues resulted in lower recall than the specific picture cues (showing more resemblance with the Anderson results). One explanation for this lies in the way in which the cues were presented, i.e., using lexical versus pictorial cues for retrieving sentences. Whereas picture cues inherently cannot have a one-to-one mapping onto the linguistic stimuli and responses, word cues can, which inevitably boosted the LSA scores for general word cues but not picture cues. Another, potentially problematic, explanation lies in the possibility that participants verbally labeled the picture cues, after which these verbal labels served as retrieval cues. In this case, the general picture cues should have elicited a wider range of lexical associations than specific cues, which could have led to a lower probability of retrieving the correct sentence (cf. generation-recognition models of retrieval, Anderson & Bower, 1972; Bahrick, 1970). Such a verbalization strategy might be confounding our contention that visual specificity is represented differently during the *encoding* of linguistic information, because it would show an interaction between the visual and linguistic domain during the *retrieval* phase only.

If this were the only strategy at work, the pattern of results of Experiment 2 should be virtually identical to that of Experiment 1 except for the main effect of cue type (with general cues now being less effective than specific cues). This is not the case. The results for the picture cues show a clear effect of encoded context for both specific and general cues, whereas the results for the word cues do not. Thus, information from the encoding phase was retrieved differently for lexical cues than for pictorial ones, which suggests that verbalization at the retrieval level cannot be the sole explanation for our findings. However, using an encoding-retrieval paradigm cannot fully exclude an offline retrieval interpretation of the results. For this reason, we performed an online sentence processing experiment to pinpoint the role of visual specificity in the creation of online mental representations.

In Experiment 3, participants read biasing and broad sentences after which they immediately decided whether a detailed (specific) or outlined (general) object in a line drawing matched the object implied by the sentence. If linguistic specificity affects the specificity with which visual information is represented, participants should perform better on the specific line drawings after reading a sentence that biased towards the depicted exemplar than after reading a sentence that avoided this bias (relative to the general line drawings, cf. our previous experiments and those by Anderson et al., 1976).

Experiment 3

Method

Participants

We recruited 327 new participants through Amazon's Mechanical Turk (http://www.mturk.com). The sample had a mean age of 30 (ranging from 17-72, SD = 11) and contained 138 females (47%). All participants were residents of the USA and native speakers of English. They were compensated with $0.80 for their participation, which required approximately 15 minutes.

Stimuli

The twenty critical sentences and line drawings were identical to those of Experiment 2. We created twenty additional filler sentences with general and specific line drawings that did not match the sentences (e.g., filler sentence 'The sporting gear was stored in the depository' was paired with a line drawing of a fish). Participants were instructed to read the 40 sentences and decide whether the subsequently depicted object matched the object mentioned in the sentence. Thus, the twenty critical sentences required a 'yes' response and the twenty filler sentences required a 'no' response. Picture cue served as between subjects factor, meaning that participants viewed only the general or only the specific line drawings. This was done to ensure that both types of critical line drawings would receive 'yes' responses after reading both types of critical sentences. Like the previous experiments, context was included as a within subjects factor. The resulting four lists each consisted of 10 biasing, 10 broad, and 20 filler sentences that were paired with either general or specific line drawings. The lists were counterbalanced across subjects and the sentences were randomly presented within subjects.

Procedure

The experiment was programmed and presented in the online Qualtrics survey research suite (http://www.qualtrics.com). The experiment started with five practice trials to familiarize the participants with the procedure, after which the experimental session started. Each trial started with a fixation mark (#) that was presented on the left of the screen for 500 ms, which was immediately followed by a sentence. Participants pressed the spacebar as soon as they had read the sentence. Next, a fixation mark (#)

was presented in the center of the screen for 500 ms, immediately followed by a line drawing. Participants indicated whether the depicted object corresponded to the object from the sentence by pressing the 'L' (yes) or 'A' (no) key on the keyboard. Additionally, one fourth of the trials was followed by a comprehension question that also required a yes ('L') or a no ('A') response, to ensure that the participants were reading the entire sentence for comprehension. Finally, a blank screen was presented for 1000 ms after which the next trial started. After the task, participants were queried about their notion of the purpose of the experiment and filled out a brief demographic questionnaire.

Results

We calculated the mean accuracy on the pictures per participant per condition, see Figure 3. Participants with a low accuracy (average <.70, N= 11) were excluded from the analysis. To obtain equal sample sizes per counterbalancing list, 22 additional participants were excluded from the data (i.e., the 11, 8, and 3 last participants were removed from one of the lists). Exclusion of all of the above participants did not affect the pattern of results or any of the significance values.

A 2 (context) x 2 (picture) mixed ANOVA revealed a significant main effect of sentence type, which demonstrated a higher accuracy for pictures after reading biasing ($M = .96$) than broad ($M = .92$) sentence contexts, $F(1,290) = 40.25$, $p < .001$, $\eta^2 = .12$. There was no significant main effect of picture type, with equal accuracies for specific and general pictures (both $Ms = .94$, $F < 1$). Importantly, there was a significant context * picture interaction ($F(1,290) = 29.66$, $p < .001$, $\eta^2 = .09$), which showed that the context effect was significantly larger for specific than for general pictures. Post-hoc two-tailed paired-samples *t*-tests showed that responses to the specific pictures were more accurate for the sentences with biasing compared to broad contexts ($t\,(145) = 7.94$, $p < .001$, $r = .40$). The general pictures showed no such difference ($t\,(145) = 0.67$, $p = .50$).

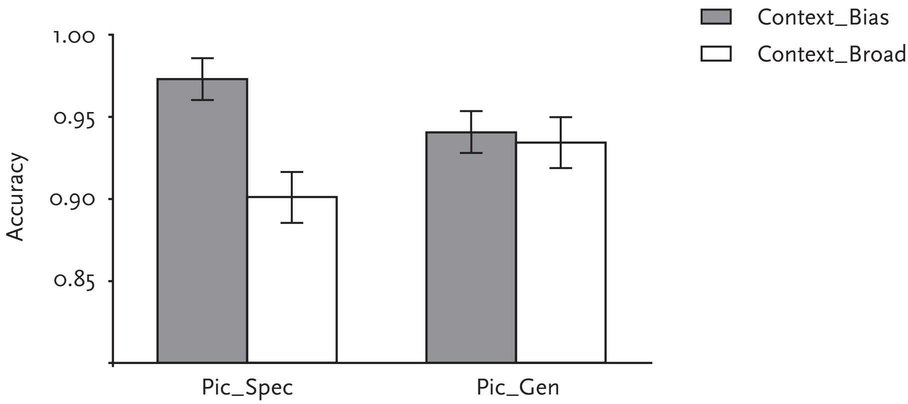


Figure 3. Accuracy on the presented pictures for biasing and broad sentence contexts. Error bars represent the 95% confidence intervals for each condition.

Next, we calculated the response times (RTs) to the correct pictures per participant per condition, see Figure 4. The response time data showed large variation and many outliers: Removing extreme response times above 4 seconds, below 200 ms, and beyond 3 standard deviations of the mean already discarded 7% of the data, more than half of which were trials from one single condition (broad context - specific picture). Therefore, we used median response times, because they are less sensitive to outliers. A 2 (context) x 2 (picture) mixed ANOVA on the median RTs showed a main effect of sentence type, with faster RTs for pictures after reading biasing (M = 985) than broad (M = 1256) sentences, $F(1,290) = 166.50$, $p < .001$, $\eta^2 = .37$. A main effect of picture type showed faster RTs for specific (M = 1070 ms) than general (M = 1171) pictures, $F(1,290) = 4.22$, $p = .04$, $\eta^2 = .01$. Finally, a significant interaction ($F(1,290) = 22.53$, $p < .001$, $\eta^2 = .07$) showed a greater context effect on specific pictures ($t(145) = 11.07$, $p < .001$, $r = .37$) than on general pictures ($t(145) = 6.76$, $p < .001$, $r = .19$).

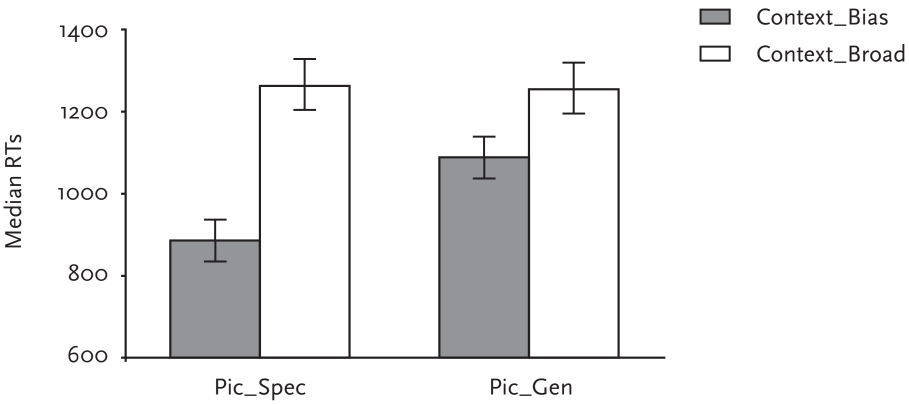


Figure 4. Median response times on the pictures for biasing and broad sentence contexts. Error bars represent the 95% confidence intervals for each condition.

Discussion

The results of Experiment 3 show that participants responded with higher speed and accuracy to pictures that were presented after biasing sentences than those that were presented after broad sentences. This demonstrates a clear processing advantage when matching visual information to biasing rather than broad sentences, which will be addressed in the general discussion. Furthermore, participants responded with higher speed and accuracy to specific than to general pictures. Finally, the advantage of biasing over broad sentence contexts was greater for the specific than the general pictures in both accuracy and speed.

In sum, these results show that participants were better able to recognize the specific line drawings as being a member of the category after reading a sentence that biased towards the depicted exemplar than after reading a sentence that avoided this bias. Thus, even though the detailed line drawing of, for example, an apple matched both the biasing and the broad sentence contexts, participants had less difficulty in responding to this specific example of fruit after reading a biasing ('The fruit forms an ingredient for cider') than a broad ('The fruit forms an ingredient for exotic salads') sentence. No such difference emerged when responding to the general outline of an apple. These findings suggest that readers activate more specific visual information when representing an implied exemplar compared to representing a category during reading, and that this activation affected performance on subsequently presented referent pictures.

General Discussion

In line with early findings, Experiment 1 showed that specific words referring to an exemplar of a category provide better retrieval cues for sentences that bias the category towards this exemplar than sentences that avoid this bias. This supports the notion that people instantiate exemplars when being confronted with general terms, at least whenever the context biases towards an instantiation. In Experiment 2 we extended these findings to the visual domain, by showing that detailed line drawings provide better retrieval cues for contexts that bias towards the instantiation than contexts that do not (to a greater degree than schematic line drawings). These findings demonstrate that conceptualizing exemplars is associated with greater visual specificity than conceptualizing the general terms themselves. In Experiment 3 we explored the nature of the interaction between linguistic and visual detail by assessing whether this effect emerged from the encoding phase of the experiment, rather than from verbalization during retrieval. In an online language processing experiment, we found that people were better capable (faster and more accurate) of recognizing detailed pictures as being a member of a category when reading biasing than broad sentence contexts. This suggests that the online instantiation of general terms is accompanied by greater visual detail than conceptualizing the general terms themselves. In other words, this shows that the visual component of the mental representation that is created during encoding of a text differs as a function of the specificity that is implied by the text. Not only do these findings support and generalize insights from language research across domains, they also provide novel insights in a grounded view on language processing.

Furthermore, the experiments showed an overall advantage of processing and retrieving biasing sentences over broad sentences. This allows for further speculations. Language naturally consists of different hierarchical levels of abstraction, for example, a superordinate level (musical instrument), a basic level (guitar), and a subordinate level (electric guitar) (Langacker, 1993; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). On the other hand, real-life physical encounters with objects naturally occur at the lowest level with specific exemplars. That is, people never interact solely with the superordinate category (musical instrument), but always with a particular exemplar (e.g., a particular electric guitar such as their father's Fender Stratocaster Sunburst). Thus, the only actual visual experience that can be stored with categorical concepts in memory occurred when interacting with exemplars at a rather specific level of visual detail. Given that (a) using language for comprehension involves mental representations of what is described, (b) such mental representations recruit visual information from memory, and (c) visual information that is stored during interactions with the environment always involves exemplars, it logically follows that people are better capable of representing and remembering sentences that bias towards exemplars than sentences that do not. The fact that this benefit was greater for specific than schematic pictures can be taken to suggest that the instantiation of a general term during the processing (encoding) of biasing contexts is already accompanied by more detailed rather than schematic visual information.

With these studies, we have taken a first step in assessing how linguistic and visual specificity interact. The findings suggest that the specificity with which visual information is represented upon processing linguistic input differs as a function of implied specificity in the text. This suggests that the detail of visual representations reflects the hierarchy of linguistic description levels, in which categories are represented in a more schematic fashion and exemplars are represented in more visual detail. This grounded interpretation of the findings is in line with Langacker's notion that the granularity of description levels may map onto that of their referent visual experiences (1999), because it suggests that *rattlesnake* is represented with greater visual detail than *snake* or *creature*, for example. These findings extend both the encoding-specificity principle and a grounded cognition account, because they support the notion that different degrees of visual specificity accompany different degrees of linguistic specificity.

Acknowledgements

We thank Bruno Bocanegra, Diane Pecher and the members of her Memory Lab, as well as the members of the Language and Cognition Lab (all at Erasmus University Rotterdam) for their constructive feedback.

Chapter 6

Summary and Discussion

We have learned that conceptual knowledge consists of information from different mental 'faculties', such as language, perception, memory, and action. Prior research and the research from this thesis show that these processes interact, meaning that they influence each other in several ways. We have specifically studied how linguistic and visual processes interact in perception and memory. In the first part of this thesis, we have used the fact that people activate visual information when processing language to gain insights in how bilinguals organize linguistic information. We have placed them in a situated setting in which we used their eye movements to relevant visual objects to examine whether they were able to 'switch off' the language they did not use. In the second part, we have asked questions about the content of visual representations. Here, we assessed whether and how language (and thereby the situation it describes) can affect the visual quality of these representations. The main findings of these studies are summarized and discussed below.

The interplay between languages in a visual world

Consider processing information in your second language, either through a conversation, through notes through a slot in the door, or by reading Matilda. You can imagine that retrieving information from your native language may help you understand your second language, but also that this extra native language knowledge may hinder you to efficiently find the right meaning of the words in the second language. Research has shown that people access the language they are not using when they process words in isolation (e.g., Dijkstra, Grainger, & Van Heuven, 1999). Some findings suggest that this native language effect diminishes in a sentence context (Schwartz & Kroll, 2006; Van Hell & De Groot, 2008), others suggest that it merely decreases (Chambers & Cooke, 2009; Libben & Titone, 2009). In Chapters 2 and 3, we examined whether bilinguals access information from both languages when listening to sentences, or whether they are able to switch off the language they are not currently using.

In **Chapter 2**, we examined whether cross-language activation occurs across word classes in a sentence context. Dutch-English bilinguals were auditorily presented with English sentences (e.g., 'The police officer spoke to the child') while looking at a visual world. The sentences contained words from distinct lexical categories with form overlap across languages (e.g., the English verb *spoke*, which overlaps phonologically with the Dutch noun for ghost, *spook*). Eye movement recordings showed that, despite the fact that the context biased towards the English verb meaning of the interlingual homophones, bilinguals directed more visual attention to depictions of the Dutch noun meaning than to unrelated distractor pictures. That this cross-language effect occurred across word classes *despite* the structural information provided by the sentence context suggests that people are broadly affected by their native language (even when it is not helping them understand what is being communicated).

In **Chapter 3** we took a different approach in studying this question. We took a step back to critically consider how eye movements can reveal interacting cognitive processes such as cross-language activation. We reviewed the literature, showing that researchers who use the visual world paradigm to detect such interactions mostly analyze the viewing tendency of participants over trials. However, we argue that they should ideally be analyzed *within individuals* and *within trials*, because this is where the interaction would take place. We performed a visual world experiment in which Dutch-English bilinguals were instructed in English to e.g., 'click on the *mice*' (with *mais* meaning corn in Dutch) in the presence of a target picture (of the English meaning of *mice*), a competitor picture (of the Dutch meaning of *mais*, being corn), and two distractor pictures. We argue that if both meanings of the word *mice* are activated in this trial, this should be reflected in eye movements to both pictures. Furthermore, the eye movements to these critical pictures should occur in close temporal relation to one another in order to conclude that both representations were interacting. We illustrate that oft-used methods have a discrepancy between the description level (within individuals and trials) and the level of analysis (averaging across individuals or trials), which may result in false acceptance or rejection of interaction hypotheses. Our solution to this problem lies in assessing eye movement transitions between critical pictures within participants and within trials. We hypothesized that co-activation should manifest through more eye movements from the English to the Dutch depiction than from the English picture to an irrelevant picture. After performing this precise test of co-activation the results indeed showed this pattern, which again confirms that bilinguals activate the meaning of both languages in a visual world setting.

To summarize, these two chapters in which bilinguals processed sentences from their second language demonstrate that (a) visually presenting the word meanings from both languages resulted in temporal dependencies between fixating these depictions, and (b) visually presenting the native language noun meaning of a second language verb resulted in more fixations to the Dutch depiction than another irrelevant picture. The findings from both chapters support and extend the existing literature in several ways. First, they suggest that people are influenced by the meaning of the language they are not using, even if the sentence context biases them in the direction of the intended meaning. This is in concordance with previous behavioral, eye tracking, and ERP studies that suggest that cross-language activation is persistent in bilinguals. Second, the findings suggest that this influence from the other language is transient, which means that it is resolved relatively quickly. This can explain why some studies find no evidence for cross-language activation in a context and some do, given that the findings in support of cross-language activation were obtained relatively early in a trial (Chambers & Cooke, 2009; Libben & Titone, 2009) and those against it relatively late (Schwartz & Kroll, 2006; Van Hell & De Groot, 2008). Third, we have specified the definition of cross-language activation (and any type of momentary cognitive interactions for that matter), which makes it possible to perform more suitable tests of cross-language activation. Fourth, we have provided a method for performing such tests, with which we have made a stronger case for interacting representations in the presence of a relevant visual world.

The question remains to what extent these lab findings generalize to more natural language processing. Reading or listening to an entire story of course provides more context than a mere sentence does. This might constrain the interpretation of a word's meaning to such an extent that any native language influence is diminished. In other words, reading Matilda in your second language might provide you with sufficient information to no longer be affected by your native language. On the other hand, you might not as easily be absorbed when processing a story in your second language, because your native language is the more dominant one and keeps competing with the words you read. This is a potentially interesting question to study in future research. A related but different approach would be to study the type of visual representations bilinguals create when processing their second language in the absence of a visual world paradigm, and the extent to which they activate this visual information. The rationale behind this is as follows. In unbalanced bilinguals who have acquired their second language at a later age, a second language can be either acquired though experience in a different language environment (immersion, connecting second language words to objects in a different environment), experience in the classroom (translation, connecting second language words to first language words), or a combination of both. The underlying mechanism for comprehending the second language might be shaped accordingly, in which case there may be large individual differences depending on the way in which a second language was learned (but perhaps also the degree of proficiency of a second language user). For example, here we have studied unbalanced bilinguals who were proficient but not fluent in their second language. In this population, the finding that one's dominant language cannot be switched off during second language processing might suggest that second language information is in some way (partially) 'grounded' in the native language. This could affect the way in which information is represented and the extent to which visual information is activated during second language processing. In one extreme case, one's second language may be largely shaped by the native language, in which case bilinguals might make similar visual representations in both languages, but slower or to a lesser extent in their second language. In the other extreme case, one's second language may be mainly shaped by experience with the second language environment, in which case bilinguals may even visually represent the world differently when using different languages. Of course, these are mere speculations that remain open to examination.

The interplay between language and the quality of visual representations

You, just like me, likely have the experience that a good story can make you vividly envision the described situation. Prior research has demonstrated that people indeed access visual information when they are processing language (e.g., Fincher-Kiefer, 2001; Horton & Rapp, 2003). That is, people represent rather specific visual features such as the shape or orientation of objects and the direction of actions (e.g., Richardson, Spivey, Barsalou, & McRae, 2003; Stanfield & Zwaan, 2001). But given that one makes such visual representations, can some mental images indeed be more vivid and detailed than others? And if so, can language cause these differences? In Chapters 4 and 5 we examined this issue.

In **Chapter 4**, we examined whether language affects the vividness of a visual representation in memory. Participants studied a picture, read a story about the depicted object, and then selected out of two pictures the one whose transparency level most resembled that of the previously presented picture. The stories contained two linguistic manipulations that have been demonstrated to affect concept availability in memory, namely object presence and goal-relevance. The results of two experiments showed that described absence of an object caused people to select the most transparent picture more often than described presence of the object. Thus, reading that there were no crayons in Matilda's classroom would result in a less vivid visual representation for crayons than reading that there were crayons. This implies that the entities that are not present in a described situation have a decreased perceivability in the referent representation. Goal relevance did not have any effect on the transparency of the selected pictures. For example, reading that Matilda did not need versus really needed the crayons to scare Miss Trunchbull would not affect the transparency of the selected pictures. This suggests that the current experiment actually tapped into visual properties of a mental representation (which can only be affected by presence, because relevance does not affect the physical appearance of a visual scene) rather than overall concept- or lexical availability (which can be affected by both presence and relevance). These findings show that the perceptual vividness of concepts in memory can differ, depending on the situation that is described by language.

In **Chapter 5** we asked a related question, namely about the specificity with which linguistic and visual information interact. People represent general category information (e.g., 'fish') in terms of specific exemplars that are implied by the context (e.g., 'shark') (Anderson et al., 1976). We used this insight to study whether different levels of implied linguistic specificity leave their traces in the visual domain. We hypothesized that instantiating exemplars for general terms should be associated with more detailed visual information than conceptualizing the general terms themselves. Experiment 1 replicated the early findings by showing that specific lexical cues ('apple') result in better retrieval for sentences biasing towards that instantiation ('The fruit forms an ingredient for cider') relative to sentences that avoid this bias ('The fruit forms an ingredient for exotic salads'), when comparing them to general lexical cues ('fruit'). Experiment 2 extended these findings to the visual domain by using detailed versus schematic line drawings as retrieval cues. The results showed that detailed line drawings provided better cues for biasing contexts than broad contexts (to a greater degree than schematic line drawings). These findings demonstrate that visual specificity interacts with linguistic specificity. In Experiment 3 we examined the nature of this interaction by assessing whether it takes place during language processing (encoding) rather than after (retrieval). In an online language processing experiment, people were faster and more accurate in recognizing detailed pictures as being a member of a category after reading biasing than broad sentence contexts. This suggests that the specificity with which visual information is activated upon encoding linguistic input differs as a function of implied specificity in the text. The results of this batch of experiments suggest that the representation of visual information reflects the hierarchy in linguistic description levels, in which categories (such as fruit) are visually represented in a more schematic fashion and exemplars (such as apple) are represented in more visual detail.

These two chapters demonstrate that (a) the perceptual properties of a described situation affect the vividness of visual information that is stored in memory, and (b) implied linguistic specificity affects the detail with which visual information is represented. These studies contribute to the literature in several ways. First, they combine early insights from language research (e.g., with respect to representing described visibility, goal-relevance, categories, and exemplars) with recent insights from grounded cognition. Their findings generalize insights from language research across domains, i.e., from the linguistic to the visual domain. Second, the results specifically show that the reactivation of visual and linguistic experiential traces that were laid down earlier during the experimental procedure were affected by manipulations in a different domain (linguistic and visual, respectively). For example, linguistic information was able to differentially reactivate previously laid down visual information, which affected the availability of this stored information. Third, we measured differences in the availability of visual information by examining the vividness or detail with which information was retrieved. This was based on the assumption that information that is more readily available should also be represented more vividly or in more detail. This approach extends behavioral evidence in support of a grounded cognition view by looking beyond opposing visual object characteristics (such as orientation) and showing that even the quality of visual representations can differ as a function of information in a different domain.

As always, the question remains to what extent these findings generalize to natural language processing. In these latter studies, we have taken a first step to assess the quality with which information is represented when retrieving it from memory. It is still open to examination whether vividness already differs during the *creation* of representations, that is, during on-line language processing. This would answer the question whether my vivid memory of Matilda originates from my vivid visual representations while reading the book, or that this is a function of vividness (in terms of activation) in my memory system. A related interesting avenue of research would be to further examine the specificity with which we represent visual information. We have found that activating 'apple' is associated with more visual detail than activating 'fruit'. But does this differentiation occur across all language levels? For example, 'the car' has a different connotation than 'your car', or even 'a car'. Do we always represent visual information to the greatest degree of specificity and vividness possible, or is our cognitive system more economic and do we only recruit the components that are necessary for comprehension? These are some potential follow-up questions that remain to be answered.

Conclusion

Reading a good book or listening to a passionate plea can draw you away from reality. It can make you experience a different reality in which you envision a different world. This phenomenon already hints at a role for visual information being involved in language processing. In this thesis, we have examined the mechanism that is at the foundation of this experience, by studying how visual and linguistic information interact. We have found that bilinguals direct their attention to visual information about the language they are not using. We have further demonstrated that the detail and vividness with which visual information is represented differs as a function of the situation that is described by language. Together with (and extending) the existing literature, these results support the notion that mental representations contain visual information. This information is laid down during previous experience with the world and is reactivated during language processing. These insights suggest that language users to some extent actually do perceive the wondrous world that is described in a story, by reactivating their own earlier perceptions from memory as a story unfolds.

References

Agresti, A. (2002). Categorical Data Analysis 2nd edition. New York: Wiley.

Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1459-1468.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 997-1010.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm'. *Cognition, 93*, B79-B87.

Altmann, G.T.M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137*(2), 190-200.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 347-386). New York: Psychology Press.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review; Psychological Review, 79*(2), 97-123.

Anderson, R. C., & McGaw, B. (1973). On the representation of meanings of general terms. *Journal of Experimental Psychology, 101*(2), 301-306.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles--A problem of polysemy. *Cognitive Psychology, 7*(2), 167-180.

Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. R. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 667-679.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database Philadelphia: University of Pennsylvania.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review, 77*(3), 215-222.

Barclay, J., Bransford, J. D., Franks, J. J., McCarrell, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13*(4), 471-481.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Barr, D. J., Gann, T.M., & Pierce, R.S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica, 137*, 201-207.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(04), 577-660.

Bishop, Y. M. M., Fienberg, S.E., & Holland, P.W. (1975). Discrete multivariate analysis: Theory and Practice. Cambridge, MA.: MIT Press.

Bogartz, R. S., & Staub, A. (2012). Gaze step distributions reflect fixations and saccades: A comment on Stephen and Mirman (2010). *Cognition, 123*(2), 325-334.

Borghi, A. M., & Riggio, L. (2009). Sentence comprehension and simulation of object temporary, canonical and stable affordances. *Brain Research, 1253*(0), 117-128.

Bower, G., & Morrow, D. (1990). Mental models in narrative comprehension. *Science, 247*(4938), 44-48.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193-209.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives: Pronouns modulate perspective-taking during narrative comprehension. *Psychological Science, 20*, 27–32.

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 137–148.

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society, 13*(4), 212-214.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1029-1040.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*(3), 476-485.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84-107.

Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11*(3), 283-294.

Cristino, F., Mathot, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods, 42* (3), 692-700.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychological Bulletin & Review, 12*(3), 453-459.

Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (2007). The influence of visual processing on phonetically driven saccades in the 'visual world' paradigm. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 471-486). Oxford, England: Elsevier.

De Groot, A. M. B., Delmaar, P., & Lupker, S. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology, 53A*, 397-428.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature, 395*, 597-600.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2010). Activation of sensory-motor areas in sentence comprehension. *Cerebral Cortex, 20*(2), 468-478.

Dijkstra, A., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41*, 496-518.

Dijkstra, A., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175-197.

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *PNAS : Proceedings of the National Academy of Sciences, 107*(37), 16396-16400.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2008). Time travel through language: Temporal shifts rapidly decrease information accessibility during reading. *Psychonomic Bulletin & Review, 15*(4), 750-756.

Dopkins, S. , Klin, C., & Myers, J. L. (1993). Accessibility of information about goals during the processing of narrative texts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(1), 70-80.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663-679.

Egidi, G., & Gerrig, R. J. (2006). Readers' experiences of characters' goals and actions. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32*(6), 1322-1329.

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research, 25*, 57-70.

Ferguson, H.J., Sanford, A.J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research, 1236,* 113-125.

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition, 29*(2), 336-343.

Findlay, J. M. (2004). Eye scanning and visual search. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 135-159). New York: Psychology Press.

Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition, 31*(1), 87-99.

Frank, A., Salverda, A. P., Jaeger, T. F., and Tanenhaus, M. (2009). Analyzing multinomial timeseries data with 'state' dependencies. The 22nd CUNY Sentence Processing Conference, Davis, CA.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Garrod, S., & Sanford, A. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 77-90.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32*(2), 99-156.

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics, 38*, 981–1014.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 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.

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

Grainger, J., & Beauvillain, C. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language, 26*, 658-672.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(2), 131-149.

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

Hallet, P. E. (1986). Eye movements. In K. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 10-11, 10-112). New York: Wiley-Interscience.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*(1-3), 335-346.

Hartsuiker, R.J., Huettig, F., & Olivers, C.N. (2011). Visual search and visual world: Interactions among visual attention, language, and working memory (introduction to the special issue). *Acta Psychologica, 137*, 123-137.

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics, 38,* 1015-1032.

Henderson, J. M., Falk, R., Minut, S., Dyer, F. C., & Mahadevan, S. (2000). Gaze control for face learning and recognition by humans and machines. *Michigan State University Eye Movement Laboratory Technical Report, 4*, 1-14.

Henderson, J. M., & Ferreira, F. (2004). Scene Perception for Psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 1-58). New York: Psychology Press.

Hirschfeld, G., & Zwitserlood, P. (2011). How vision is shaped by language comprehension — top-down feedback based on low-spatial frequencies. *Brain Research, 1377*(0), 78-83.

Holmes, K.J., & Wolff, P. (2011). Simulating realism in language comprehension. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp.2884-2889). Austin, TX: Cognitive Science Society.

Holt, L. E., & Beilock, S. L. (2006). Expertise and its embodiment: Examining the impact of sensorimotor skill expertise on the representation of action-related text. *Psychonomic Bulletin & Review, 13*(4), 694-701.

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review, 10*(1), 104-110.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96*, B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language, 57*, 460–482.

Hwang, A., Wang, H., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research, 51*. 1192-1205.

Jackendoff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition, 26*(2), 89-114.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAS (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446.

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

Just, M. A., & Carpenter, P. A. (1978). Inference processes during reading: Reflections from eye fixations. In J. M. Senders, D. F. Fisher & R. A. Monty (Eds.), *Eye movements and the higher psychological functions*. New Jersey: Lawrence Erlbaum.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition, 94*(3), B79-B89.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38*(7), 1033.

Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind.* Cambridge University 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*(2), 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259-284.

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10*(1), 1-40.

Langacker, R. W. (1993). Grammatical traces of some 'invisible' semantic constructs. *Language Sciences, 15*(4), 323-355.

Li, P., & Yip, M. C. (1998). Context effects and the processing of spoken homophones. *Reading and Writing: An Interdisciplinary Journal, 10*, 223-243.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 381-390.

Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian Hidden Markov Model. *Psychometrika, 68* (4), 519-541.

Lincoln, A. E., Long, D. L., & Baynes, K. (2007). Hemispheric differences in the activation of perceptual information during sentence comprehension. *Neuropsychologia, 45*(2), 397.

Lupyan, G., & Thompson-Schill, S.L. (2011). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, No Pagination Specified. doi: 10.1037/a0024904.

Lutz, M. F., & Radvansky, G. A. (1997). The fate of completed goal information in narrative comprehension. *Journal of Memory and Language, 36*(2), 293-310.

MacDonald, M.C., & Just, M.A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 633-642.

Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior, 10*, 480-487.

Magliano, J. P., Taylor, H. A., & Kim, H. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory & Cognition, 33*(8), 1357-1367.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 1-24.

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2007). Sensitivity to phonological similarity within and across languages. *Journal of Psycholinguistic Research, 37*, 141-170.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940-967.

Marian, V., & Spivey, M. J. (2003). Competing activation in bilingual language processing: Within- and between- language competition. *Bilingualism: Language and Cognition, 6*(2), 97-115.

Marian, V., Spivey, M. J., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eye tracking and brain imaging. *Brain and Language, 86*, 70-82.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372-380.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from 'lexical' garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language,60*(1), 65-91.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*(1), 89-115.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*, 475-494.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior, 22*(5), 526-534.

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology, 62*(6), 1108-1114.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science, 14*(2), 119-124.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 785-794.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*(02), 253-336.

Rapp, D. N., & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*(5), 988-1001.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618-660.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology, 102*(4), 648-656.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*(5), 767-780.

Risen, J. L., & Critcher, C. R. (2011). Visceral fit: While in a visceral state, associated states of the world seem more likely. *Journal of Personality and Social Psychology, 100*(5), 777-793.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382-439.

Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1155-1178.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*, 197-212.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(03), 417-424.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic Access of the Meanings of Ambiguous Words in Context: Some Limitations of Knowledge-Based Processing. *Cognitive Psychology, 14*, 489-537.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia, 45*(12), 2802-2810.

Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research, 9*, 237-251.

Spivey-Knowlton, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (1998). Integration of visuospatial and linguistic information: Language-comprehension in real-time and real-space. In P. Olivier & K. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 201-214). Hillsdale, NJ: Erlbaum.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281-284.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153-156.

Stephen, D. G., & Mirman, D. (2010). Interactions dominate the dynamics of visual cognition. *Cognition, 115*(1), 154-165.

Suh, S. Y., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language, 32*(3), 279-300.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing. *Studies in Second Language Acquisition 28*(3), 387-422.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 645-659.

Swinney, D. A., & Hakes, D. T. (1976). Effects of prior context upon lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 681-689.

Tabossi, P., & Johnson-Laird, P. (1980). Linguistic context and the priming of semantic information. *The Quarterly Journal of Experimental Psychology, 32*(4), 595-603.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*(1), 49-100.

Tanenhaus, M., Frank, A., Jaeger, T. F., Masharov, M., and Salverda, A.P. (2008). The art of the state: Mixed effect regression modeling in the Visual World. The 21st CUNY Sentence Processing Conference, Chapel Hill, NC.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*, 427-440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77-99.

Treisman, A., M., & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology, 12*, 97-136.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica, 116*(2), 185-203.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, *20(8)*, 923-927.

Vandeberg, L., Eerland, A., & Zwaan, R. A. (2012). Out of mind, out of sight: Language affects perceptual vividness in memory. *PLOS ONE, 7*(4), e36154.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension*. New York: New York: Academic Press.

Van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica, 128*, 431-451.

Van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*, 2706-2716.

Van Schie, H. T., Mars, R. B., Coles, M. G., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience, 7,* 549–554.

Vermunt, J. K., Langeheine R., & Böckenholt, U. (1999). Latent class Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, *24*, 178–205.

Vermunt, J.K., & Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced.* Belmont MA: Statistical Innovations Inc.

Vermunt, J.K., & Magidson, J. (2008). *LG-Syntax user's guide: Manual for Latent GOLD 4.5 Syntax module*. Belmont MA: Statistical Innovations Inc.

Von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language, 65*, 109-127.

Von Studnitz, R. E., & Green, D. W. (2002). Interlingual homograph interference in German-English bilingual: Its modulation and locus of control. *Bilingualism: Language and Cognition, 5*, 1-23.

Williams, L. G. (1966). The effect of target specification on objects fixated during visual search. *Perception and Psychophysics, 1*, 315-318.

Wittgenstein, L. (1953). *Philosophical Investigations.*

Wolfe, J. M. (1994). Guided Search 2.0. A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

Wolter, L., Gorman, K. S., & Tanenhaus, M. K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of Memory and Language,65*(3), 299-317.

Yaxley, R.H., & Zwaan, R.A. (2007). Simulating visibility during language comprehension. *Cognition*, *105,* 229-236.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 1-14.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science : A Journal of the American Psychological Society, 8*(1), 15-18.

Zwaan, R. A. (2003). The immersed experiencer: Toward an embodied theory of language comprehension. *The Psychology of Learning and Motivation, 44*, 35-62.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292-297.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28*(4), 611-619.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171.

Samenvatting

‘Verbeelding’ in taal: Visuele representaties en taalverwerking

Mijn favoriete kinderboek was (en is nog steeds) Matilda, geschreven door Roald Dahl. Het boek gaat over Matilda Wurmhout, een bijzonder en leergierig vijfjarig meisje. Helaas heeft ze geen fijne ouders die haar maar een vervelende betweter vinden. Zij dwingen haar om magnetronmaaltijden te eten en naar stomme spelprogramma's op tv te kijken, terwijl Matilda liever leest. Het hoofd van Matilda's school, mevrouw Bulstronk, zorgt ook al niet voor een stimulerende omgeving. Zij heeft een hekel aan kinderen en laat dit zien door ze op te sluiten in het Stikhok (een smalle kast met spijkers en glasscherven aan de binnenkant) of ze aan hun haar over het schoolplein te slingeren. Op een dag ontdekt Matilda dat ze telekinetische krachten heeft waarmee ze voorwerpen kan laten bewegen. Ze besluit deze te gebruiken om streken uit te halen en haar ouders en mevrouw Bulstronk een lesje te leren. Het magische van dit boek was dat ik de realiteit om me heen vergat en in Matilda's wereld werd getrokken: ik kon haar nieuwsgierigheid en frustratie voelen, ik stelde me haar vaders gezicht voor toen hij erachter kwam dat ze zijn hoed met superlijm op zijn hoofd had geplakt, en ik zag levendig voor me hoe ze haar kracht gebruikte om krijtjes te laten vliegen en zinnen op het schoolbord te laten schrijven om mevrouw Bulstronk de stuipen op het lijf te jagen. Hoe is dit mogelijk? Hoe kunnen abstracte symbolen zoals letters en woorden op een pagina tot leven komen, je aangrijpen, levendige beelden oproepen, en ervoor zorgen dat je de beschreven gebeurtenissen bijna zelf lijkt te beleven?

Mentale representaties

Het *mentaal representeren* (voorstellen) van de echte of een denkbeeldige wereld ligt aan de basis van het menselijk denken, leren en functioneren. Daarom hebben mentale representaties altijd veel aandacht gehad in de wetenschap, bijvoorbeeld in de cognitieve psychologie, filosofie, en kunstmatige intelligentie. Het begrijpen van een verhaal (of ieder soort geschreven of gesproken taal) vereist het maken van een mentale representatie, ook wel een mentaal model genoemd. Zulke mentale modellen moeten niet enkel de tekst zelf representeren, maar vooral hetgeen waar de tekst over gaat (Johnson-Laird, 1983; Sanford & Garrod, 1981; van Dijk & Kintsch, 1983). Om Roald Dahl's verhaal te begrijpen en erin mee te gaan is het niet genoeg om alleen de woorden die hij opschreef te representeren, maar juist de *situatie* die hij beschreef. Veel empirisch onderzoek laat zien dat mensen inderdaad zulke *situatiemodellen* maken. Beschouw bijvoorbeeld de twee zins-paren uit een studie van Bransford, Barclay, en Franks (1972):

Three turtles rested on a floating log, and a fish swam beneath them. (1a)
Three turtles rested on a floating log, and a fish swam beneath it. (1b)

Three turtles rested beside a floating log and a fish swam beneath them. (2a)
Three turtles rested beside a floating log and a fish swam beneath it. (2b)

Beide zins-paren bevatten dezelfde woorden en hebben dezelfde structuur. Echter, ze beschrijven verschillende situaties: Paar 1 beschrijft een identieke opstelling (de schildpadden drijven op een boomstronk waaronder een vis zwemt), maar paar 2 beschrijft een andere opstelling (2a: de vis zwemt onder de schildpadden en de boomstronk drijft daarnaast, 2b: de vis zwemt onder de boomstronk en de schildpadden zwemmen daarnaast). De proefpersonen in deze studie kregen een lijst met zinnen te lezen waarin één zin van een paar stond (bijvoorbeeld zin 1a). Daarna kregen ze beide zinnen uit een paar te zien en moesten ze beslissen welke zin ze eerder gelezen hadden. Ze vonden het relatief gemakkelijk om deze beslissing te maken voor de zinnen in paar 2 (die een andere opstelling beschreven), maar erg moeilijk voor de zinnen in paar 1 (die dezelfde opstelling beschreven). Dit toont aan dat het geheugen voor zinnen georganiseerd is op basis van de beschreven situatie (bijvoorbeeld een bepaalde opstelling) en niet enkel op basis van de woorden in de tekst zelf.

Later onderzoek heeft aangetoond dat mensen nog veel meer informatie uit een beschreven situatie representeren, bijvoorbeeld over de personages, objecten, locaties, acties, doelen, intenties, tijd, en perspectieven die beschreven zijn in een tekst (bv. Brunyé et al., 2009; Ditman, Holcomb, & Kuperberg, 2008; Egidi & Gerrig, 2006; Glenberg, Meyer, & Lindem, 1987; Magliano, Taylor, & Kim, 2005; Rapp & Taylor, 2004; Zwaan, 1999; Zwaan, Langston, & Graesser, 1995). Echter, weten *wat* we representeren bij het begrijpen van taal zegt nog niets over *hoe* we deze informatie representeren. Hoe representeerde ik de personages en objecten uit het verhaal van Matilda? Hoe kunnen mensen überhaupt betekenis geven aan de concepten die ze representeren?

Betekenis

Om betekenis te kunnen geven aan taal moeten woorden op een of andere manier verbonden worden aan datgene waar ze naar refereren (Harnad, 1990). Wanneer dit niet het geval was zouden woorden enkel willekeurige formele symbolen zijn. Laten we het krijtje uit Matilda als voorbeeld nemen. Het feit dat er naar dit object verwezen wordt als *krijtje* door Nederlanders, maar *crayon* door Engelsen en *tiza* door Spanjaarden toont al aan dat de taal die we gebruiken om aan de wereld te refereren arbitrair is. Om het woord *krijtje* (of *crayon* of *tiza*) te begrijpen moeten we het op een bepaalde manier koppelen aan het daadwerkelijke object. Dit wordt mooi duidelijk gemaakt in het volgende gedachte-experiment dat werd bedacht door Searle (1980). Stel je voor dat je in een gesloten kamer zit met een boek, papier en potlood. Door een brievenbus in de deur krijg je briefjes met Chinese tekens toegeschoven. Je spreekt geen Chinees, maar in het boek staan instructies die uitleggen welke Chinese symbolen je zelf op moet schrijven in reactie op de symbolen op het briefje. Je hebt dus geen woordenboek, maar een boek met enkel Chinese tekens die jij over kunt schrijven. Je schrijft jouw reactie, stopt deze in de brievenbus, en ontvangt weer een nieuw briefje. Dit gaat een tijdje zo door. Stel je voor dat je deze taak zo goed uitvoert dat je degene aan de andere kant van de deur, iemand met Chinees als moedertaal, overtuigt dat jij zelf een Chinese moedertaalspreker bent. Echter, dit ben je natuurlijk niet. Je *begreep* helemaal niets van wat je schreef. Stel je nu voor dat je een woordenboek hebt dat de Chinese symbolen naar je moedertaal, het Nederlands, kan vertalen. Na deze vertaling zou je natuurlijk wel gaan begrijpen wat je hebt gelezen en opgeschreven. Echter, we hebben zojuist gezien in het voorbeeld met de krijtjes dat de symbolen uit je moedertaal (Nederlandse woorden) net zo abstract en willekeurig zijn als Chinese (en Engelse en Spaanse) symbolen. Hoe kan het dan dat je Nederlandse symbolen wel begrijpt? Waar kun je deze symbolen naar 'vertalen' zodat er begrip plaatsvindt? Of nog specifieker, hoe kan een Nederlandse uitgave van Matilda mij een andere wereld laten ervaren terwijl een Chinese uitgave mij enkele rare krabbeltjes op een pagina laat ervaren? Het antwoord is: Door deze abstracte symbolen te verankeren (gronden) in ervaring.

In 1999 publiceerde de wetenschapper Barsalou een invloedrijk artikel, 'Perceptual symbol systems' genaamd. In dit artikel beschreef hij dat er een analoge relatie moet zijn tussen een symbool (bijvoorbeeld het woord *krijtje*) en hetgeen waar dit symbool aan refereert (het eigenlijke object). Hij stelde dat kennis pas analoog (gelijkend) aan de externe wereld kan zijn als het conceptuele systeem dat deze kennis gebruikt (het menselijk brein) perceptueel van aard is. Dit komt neer op het volgende. De representaties die we maken wanneer we denken of taal verwerken moeten in zekere zin gelijk zijn aan de representaties die we maakten toen we de respectievelijke situatie zelf meemaakten. Bijvoorbeeld, betekenis geven aan het woord *krijtje* houdt in dat we de sporen in ons geheugen moeten activeren die we eerder hebben gemaakt tijdens onze ervaringen met krijtjes. Dit heeft als gevolg dat het waarnemen van de wereld om ons heen en het opslaan van deze informatie ervoor zorgt dat wij kunnen representeren hoe een krijtje er uitziet, dat het gebruikt wordt om mee te schrijven, hoe het voelt of klinkt als je ermee schrijft, dat het kan breken, dat het niet eetbaar is, dat het vaak gebruikt wordt door docenten, enzovoort. Verder stelt Barsalou dat losse componenten van een representatie gebruikt en gecombineerd kunnen worden om een nieuwe representatie te maken. Dit zorgt ervoor dat, zelfs als we geen fysieke ervaring hebben met een bepaalde situatie zoals de vliegende krijtjes die Roald Dahl beschreef, we onze ervaring met de losse componenten 'vliegend' en 'krijtjes' in kunnen zetten om de betekenis van vliegende krijtjes voor ons te zien.

Kortom, om betekenis te geven aan taal moeten mentale representaties woorden verbinden aan hetgeen ze naar verwijzen. Dit kan enkel als representaties niet alleen talig zijn, maar ook perceptuele (visueel, auditief) en functionele (motorische acties) informatie bevatten die verankerd is in eigen ervaringen met de omgeving[21]. Dit idee was niet nieuw: Eerder al hadden theoretici een rol voor visuele en motorische informatie weggelegd bij het representeren van concepten (Jackendoff, 1987; Lakoff, 1990; Langacker, 1986; Talmy, 1988). Echter, de theorie van Barsalou (1999) stelde specifiek dat taal niet een losstaande module in het brein kan vormen, maar dat taal moet *interacteren* (samenwerken) met de neurale systemen die actief zijn bij perceptie en actie. In dit geval zouden de gebieden in het brein die verantwoordelijk zijn voor het *waarnemen* van de omgeving ook tot op zekere hoogte gebruikt moeten worden bij het *representeren* van een omgeving. Met deze claim leverde Barsalou een theorie over het mechanisme dat mentale representaties produceert. Hierdoor konden onderzoeker op zoek gaan naar bewijs voor of tegen deze processen in het menselijk gedrag en het menselijk brein.

Bewijs voor visuele representaties

Inmiddels is er veel bewijs vergaard dat mensen perceptuele informatie activeren bij het begrijpen van taal. Het lijkt erop dat deze perceptuele informatie in het geheugen wordt georganiseerd op basis van modaliteiten (zintuigen) en eigenschappen (bv. Barclay, Bransford, Franks, McCarrel, & Nitsch, 1974; Pecher, Zeelenberg & Barsalou, 2003). Bijvoorbeeld, Pecher, Zeelenberg, en Barsalou (2003) lieten in hun experimenten zien dat proefpersonen sneller een eigenschap herkenden (bv. auditief: 'A blender can be loud') als deze voorafgegaan werd door een eigenschap in dezelfde modaliteit (auditief: 'Leaves can be rustling') vergeleken met een eigenschap in een andere modaliteit (visueel: 'A highway sign can be green'). Het feit dat er op dezelfde modaliteit sneller gereageerd wordt dan op een andere modaliteit suggereert dat perceptuele informatie uit verschillende modaliteiten (zicht, gehoor, smaak, geur, tast, actie) op een andere manier gerepresenteerd wordt. We gaan nu verder in op de visuele modaliteit, aangezien deze misschien licht kan werpen op mijn ervaring met het 'verbeelden' van Matilda.

21 *Deze ervaring kan van persoon tot persoon verschillen. De representatie die je maakte bij het lezen over Dahl's Matilda, Rowling's Harry Potter, Tolkien's Frodo, of Larsson's Lisbeth Salander kan best anders zijn dan het beeld dat de auteurs of de producenten van de respectievelijke films in gedachten hadden.*

De eerste studies naar het maken van visuele representaties bij het verwerken van taal lieten zien dat mensen sneller plaatjes kunnen identificeren waarin de visuele eigenschappen van een object congruent (t.o.v. incongruent) zijn met een beschreven situatie. Bijvoorbeeld, in een ouder experiment van Stanfield en Zwaan (2001) lazen proefpersonen zinnen als '*John put the pencil in the cup*'. Ze herkenden een plaatje van een potlood daarna sneller als de oriëntatie van het potlood overeenkwam met de oriëntatie die geïmpliceerd werd in de zin (verticaal potlood) dan als die niet overeenkwam (horizontaal potlood). Dit laat zien dat mensen de visuele kenmerken van objecten tijdens het lezen activeren, en dat deze activatie het herkennen van de overeenkomende plaatjes beïnvloedt. Hierna hebben meer gedragsonderzoeken laten zien dat verschillende visuele kenmerken van objecten en acties worden gerepresenteerd bij het verwerken van taal, zoals vorm, oriëntatie, kleur, grootte, en richting (Fincher-Kiefer, 2001; Horton & Rapp, 2003; Lupyan & Thompson-Schill, 2012; Richardson, Spivey, Barsalou, & McRae, 2003; Borghi & Riggio, 2009; Hirschfeld & Zwitserlood, 2010; Kaschak et al., 2005; Zwaan, Stanfield, & Yaxley, 2002). Ook is er bewijs uit fMRI (hersen-) onderzoek dat de hersengebieden die verantwoordelijk zijn voor visuele perceptie ook actief zijn als er taal over visuele situaties wordt verwerkt (Desai, Binder, Conant, & Seidenberg, 2009; Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, 2007).

Al deze bevindingen leveren bewijs dat mentale representaties verankerd zijn in perceptie, oftewel dat situatiemodellen visuele informatie bevatten. Dit kan verklaren hoe mensen het gevoel hebben een beschreven situatie waar te nemen. De bevindingen suggereren namelijk dat lezers op een bepaalde manier daadwerkelijk de wereld die is beschreven in boeken ervaren, door hun eigen percepties op te halen uit hun geheugen naarmate ze de woorden, zinnen en paragrafen uit het boek verslinden. Terwijl het verhaal zich ontvouwt kunnen de verschillende componenten van deze ervaringen samen een unieke representatie creëren van de wereld die beschreven is in het boek.

Echter, conceptuele representaties bestaan niet enkel uit visuele informatie. Ze bevatten ook nog linguïstische, semantische, en functionele kenmerken en worden gevormd door iemands wereldkennis en overtuigingen (bv. Murphy & Medin, 1985). Ondanks dat visuele informatie maar één component is van een concept, heeft de wisselwerking tussen taal en visuele informatie bewezen een erg handig middel te zijn voor het bestuderen van menselijke cognitie (menselijk denken). In het eerste deel van dit proefschrift gebruiken we deze wisselwerking om inzicht te krijgen in hoe mensen meerdere talen organiseren in het brein. In het tweede deel gebruiken we visuele informatie niet meer enkel om taal te bestuderen, maar onderzoeken we hoe taal het geheugen voor visuele informatie beïnvloedt en hoe visuele informatie het geheugen voor taal beïnvloedt.

De wisselwerking tussen talen gemeten met visuele referenten

Kennis van meerdere talen in één brein kan zorgen voor verwarrende, vreemde, of grappige situaties. Denk bijvoorbeeld aan *Mama Appelsap*, een item in een programma van radiozender 3FM. Mensen bellen wekelijks naar 3FM om in de uitzending te vertellen wat zij voor vreemde teksten in liedjes horen. Meestal luisteren deze mensen naar nummers in een andere taal, bijvoorbeeld in het Engels of Spaans, maar horen ze hier Nederlandse teksten in. Zo was er iemand die in het nummer 'Could you be loved' van Bob Marley standaard 'theezakje*'* hoorde (in plaats van '*say something'*). Ook waren er mensen die de Red Hot Chili Peppers hoorde zingen over 'Bob Marley boven in een bakfiets' (i.p.v. *'Bob Marley, poet and a prophet'*), die in een nummer van de Wee Papa Girl Rappers 'Peter heeft het overleefd' hoorden (i.p.v. *'Save the nation, don't delay'*), die de Spaanstalige Shakira 'Ik heb trek in peren' hoorden zingen, of die 'Ik heb een fietsenrek' hoorden in het Italiaanse gezang van Andrea Bocelli. Dit is een van de vele voorbeelden waarin gelijkende klanken tussen talen kunnen zorgen voor grappige interpretatiefouten. Denk bijvoorbeeld ook aan Engelse woorden als *bill* (bil), *bias* (bajes), of *pet* (pet), die in het Nederlands ook bestaan maar een andere betekenis hebben, wat tot komische situaties kan leiden. Echter, zulke woorden met gelijkende klanken tussen talen, ook wel woorden met *fonologische gelijkenissen* genoemd, zijn niet alleen geschikt voor woordgrappen. Ze komen ook handig van pas bij het bestuderen hoe meerdere talen worden opgeslagen in één brein.

Veel studies hebben gelijkenissen tussen talen gebruikt om te onderzoeken hoe ons cognitieve systeem omgaat met kennis van meerdere talen. Ze bestudeerden hoe tweetaligen toegang krijgen tot hun lexicale (woord) informatie. Het merendeel van deze studies heeft gemeten hoe snel tweetaligen losse woorden herkennen met en zonder gelijkenissen tussen talen (bv. Brysbaert, Van Dyck, & Van de Poel, 1999; Caramazza & Brones, 1979; Grainger & Beauvillain, 1987; Macnamara & Kushnir, 1971; Schulpen, Dijkstra, Schriefers, & Hasper, 2003; von Studnitz & Green, 2002). De resultaten van deze studies laten zien dat tweetalingen woorden met overeenkomsten tussen talen anders verwerken dan woorden zonder deze overeenkomsten. Neem bijvoorbeeld het Engelse woord *note* (briefje, notitie) wat fonologische gelijkenis heeft met het Nederlandse woord *noot*. Nederlanders met als tweede taal Engels hebben meer tijd nodig om het Engelse woord *note* te herkennen dan een Engels woord dat geen gelijkenis heeft met het Nederlands (Dijkstra, Grainger, & Van Heuven, 1999). Dit wijst erop dat tweetaligen moeite hebben met het beslissen welke *note* er bedoeld wordt (de Nederlandse of de Engelse betekenis). De onderzoekers schrijven dit toe aan invloeden van de andere taal. Met andere woorden, het lijkt erop dat tweetaligen die woorden uit de ene taal verwerken, hun kennis van de andere taal niet uit kunnen zetten. Dit onvermogen om de andere taal uit te zetten wordt ook wel *cross-linguïstische activatie* genoemd.

Er is sinds deze eerste studies steeds meer onderzoek gedaan naar cross-linguïstische activatie met verschillende soorten experimenten, waaronder experimenten die gebruik maken van het '*visuele wereld paradigma*' (Cooper, 1974). Dit paradigma kan inzicht geven in wat er tijdens het verwerken van taal gebeurt, en wordt gebruikt om te onderzoeken wanneer en in welke mate informatie uit de andere taal actief is. In dit paradigma worden oogbewegingen van mensen gemeten terwijl ze naar zinnen luisteren en naar *referenten* (plaatjes op een scherm) kijken. Omdat taal ervoor zorgt dat visuele informatie in het brein geactiveerd wordt, kunnen oogbewegingen in een relevante visuele context de snelle mentale processen meten die plaatsvinden tijdens het verwerken van taal (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

Het visuele wereld paradigma kan worden gebruikt om cross-linguïstische activatie te meten van woorden in een zinscontext, en gaat daarmee verder dan meten wat er gebeurt bij het verwerken van losse woorden. Omdat woorden in de werkelijkheid bijna nooit geïsoleerd voorkomen maar meestal verwerkt worden in een zin, is onderzoek met zinnen en verhalen nodig om de bevindingen te kunnen generaliseren naar meer natuurlijk taalgebruik. Laten we nog even teruggaan naar het Engelse woord *note* (noot). Onderzoek met losse woorden liet zien dat tweetaligen moeite hebben met het beslissen welke *note* er bedoeld wordt (de Nederlandse of de Engelse betekenis). Echter, het horen van 'I wrote you a *note*' kan deze moeite verminderen, omdat de context informatie geeft over de bedoelde taal (Engels) en de bedoelde betekenis (iets dat je kunt schrijven). In **Hoofdstuk 2** onderzoeken we cross-linguïstische activatie in een zinscontext. Deze context bevat niet enkel informatie over de bedoelde taal en betekenis van een woord, maar ook over de woordklasse. Tweetaligen met als moedertaal Nederlands en later aangeleerde taal Engels luisterden naar Engelse zinnen (bv. '*The police officer spoke to the child*') terwijl ze naar plaatjes op een scherm keken (het visuele wereld paradigma). De zinnen bevatten woorden met fonologische gelijkenissen tussen verschillende talen en woordklassen, bijvoorbeeld het Engelse werkwoord *spoke* (sprak) dat lijkt op het Nederlandse zelfstandig naamwoord *spook* (fantoom). Oogbewegingen lieten zien dat tweetaligen meer aandacht hadden voor een plaatje van een spook dan voor een ander plaatje dat niets te maken had met de zin, zonder dat ze het zelf doorhadden. Met andere woorden, ze keken naar een afbeelding van het Nederlandse zelfstandig naamwoord *spook* ondanks dat de zinscontext stuurde naar de betekenis van het Engelse werkwoord *spoke*. Dit suggereert dat tweetaligen beïnvloedt worden door hun moedertaal zonder dat ze het doorhebben en zelfs wanneer dit niet helpt bij het begrijpen van wat ze horen. Kortom, mensen lijken er weinig aan te kunnen doen dat ze een Mama Appelsap ervaren.

In **Hoofdstuk 3** nemen we een andere benadering voor dezelfde soort vraag. We doen een stap terug en kijken kritisch naar hoe oogbewegingen het beste een wisselwerking tussen representaties, bijvoorbeeld bij cross-linguïstische activatie, kunnen meten. We bespreken eerst de literatuur waarin onderzoekers het visuele wereld paradigma gebruiken om wisselwerkingen (*interacties*) te meten. Hieruit blijkt dat deze onderzoekers hun analyses vaak doen over het gemiddelde kijkgedrag van proefpersonen, dat wil zeggen, het kijkgedrag dat is samengevat over trials (zinnen). Wij beargumenteren dat deze analyses niet over gemiddelden moeten gaan, maar het beste gedaan kunnen worden per proefpersoon en per zin. Dat is namelijk het niveau waarop tijdelijke cognitieve interacties zoals cross-linguïstische activatie moeten plaatsvinden. Simpel gezegd, er is geen sprake van cross-linguïstische activatie als proefpersonen bij de ene zin enkel de Nederlandse betekenis activeren en bij een andere zin enkel de Engelse betekenis. Dit komt omdat het Nederlands en Engels elkaar dan niet kunnen *beïnvloeden*. Echter, een gemiddelde over deze zinnen (waarin mensen de helft van de tijd kijken naar het Nederlands en de helft naar het Engels) zou suggereren dat proefpersonen beide talen activeerden, waaruit de conclusie zou worden getrokken dat ze elkaar wel beïnvloedden. Wij laten met dit voorbeeld zien dat er bij methoden die met gemiddelden werken onjuiste uitspraken kunnen worden gedaan over interacties en stellen daarom een methode voor die minder foutgevoelig is. Deze methode zou moeten laten zien dat tweetaligen beide betekenissen bij één zin activeren om te kunnen zeggen dat de ene taal de andere beïnvloedt. We deden hiervoor weer een experiment met oogbewegingen waarin Nederlands-Engelse tweetaligen Engelse opdrachten hoorden, zoals '*Click on the mice*' (waarin het Engelse *mice* muizen betekent, wat fonologisch lijkt op het Nederlandse *mais*, de groente). Ze keken naar vier afbeeldingen, een van de Engelse betekenis (muizen), een van de Nederlandse betekenis (mais), en twee ongerelateerde plaatjes (bladeren en appels). We redeneerden dat als beide betekenissen van *mice* geactiveerd worden en elkaar beïnvloeden, beide plaatjes na elkaar bekeken moeten worden bij het luisteren naar deze zin. We hebben dit gemeten door per zin te kijken of mensen hun aandacht vaker wisselt tussen het Engelse en Nederlandse plaatje dan tussen een andere combinatie van plaatjes. Wij vonden met deze strenge test inderdaad het verwachte kijkpatroon, wat wederom bevestigt dat tweetaligen de betekenis van woorden uit beide talen activeren in de aanwezigheid van een relevante visuele wereld.

Samenvattend, deze twee hoofdstukken waarin tweetaligen zinnen in hun tweede taal verwerkten lieten zien dat (a) het visueel presenteren van de woordbetekenis in beide talen (muizen en mais) zorgde voor oogbewegingen die op en neer gingen tussen de afbeeldingen van beide betekenissen en (b) het visueel presenteren van de betekenis van het zelfstandig naamwoord uit de moedertaal (een spook) bij het luisteren naar het werkwoord uit de tweede taal (*spoke*) zorgde voor meer aandacht voor de Nederlandse afbeelding dan een ander plaatje dat ook niets met de betekenis van de zin te maken had. Deze bevindingen dragen op verschillende manieren bij aan de bestaande literatuur. Allereerst laten ze zien dat tweetaligen beïnvloedt worden door de taal die ze op dat moment niet gebruiken, zelfs als de zin ze naar de juiste betekenis stuurt. Dit is in overeenstemming met de bestaande literatuur die suggereert dat tweetaligen de taal die ze niet gebruiken niet uit kunnen schakelen. Ten tweede suggereren onze resultaten dat deze invloeden van de andere taal nog voor het einde van de zin opgelost zijn. Mensen keken na het zien van het spook bijvoorbeeld weer gewoon naar het plaatje dat wel met de zin te maken had (in het geval van '*The police officer spoke to the child*' een plaatje van een kind). Ten derde hebben we de definitie van cross-linguïstische activatie (of wat voor cognitieve wisselwerking dan ook) aangescherpt, waardoor we meer gepaste tests kunnen doen voor het meten van interacties tussen cognitieve processen. Ten vierde hebben wij een methode beschreven voor zulke tests, waarmee we sterker bewijs hebben geleverd voor een wisselwerking tussen representaties (in dit geval cross-linguïstische activatie) bij het zien van een relevante visuele wereld.

De wisselwerking tussen taal en de kwaliteit van visuele representaties

Tot nu toe hebben we mensen in een omgeving geplaatst waarin relevante visuele referenten getoond werden bij het luisteren naar zinnen. Op deze manier hebben we bestudeerd hoe tweetaligen linguïstische (taal) informatie uit twee talen organiseren. In het tweede deel van dit proefschrift gebruiken we visuele informatie niet enkel om taal te bestuderen, maar vragen we ons specifiek af hoe taal de levendigheid en het detail van visuele representaties beïnvloedt. We hebben waarschijnlijk allemaal de subjectieve ervaring dat taal de kwaliteit van representaties kan beïnvloeden, net zoals mijn ervaring tijdens het lezen van Matilda. Het lezen van een goed boek, beluisteren van een songtekst, opvangen van een geanimeerd gesprek, of ophalen van herinneringen met een oude bekende kan levendige beelden oproepen. Maar zelfs minder aangrijpend taalgebruik kan dit effect hebben. Ik herinner me de keer dat de begeleider van dit proefschrift de stimuli voor een van onze experimenten opnam. Hij zat achter een microfoon in het lab en las zinnen voor van een stuk papier. Hij was zich er van bewust dat deze zinnen fonologische gelijkenissen tussen talen bevatten en was geïnstrueerd om de zinnen op een zo neutraal mogelijke toon voor te lezen. Op een gegeven moment moest hij de zin 'De jongen bedekte zijn haar met de pet' voorlezen. Vlak voordat hij het woord *pet* uitsprak sloeg zijn cross-linguïstische activatie toe, wat hem deed beseffen dat *pet* in het Engels natuurlijk huisdier betekent. Hierdoor barstte hij in lachen uit, waarna hij uitlegde dat hij zich een jongen voorstelde dat een klein hondje op zijn hoofd zette. Dit zorgde bij mij natuurlijk voor hetzelfde beeld, waarna we samen in een lachstuip belandden en zeker 5 pogingen nodig hadden voordat de zin fatsoenlijk opgenomen was. Gelukkig hadden niet alle zinnen deze uitwerking op ons.

Dit voorbeeld illustreert dat taal soms gedetailleerde en levendige beelden op kan roepen, maar niet altijd. Eerder onderzoek liet al zien dat mensen inderdaad visuele informatie representeren bij het verwerken van taal (bv. Fincher-Kiefer, 2001; Horton & Rapp, 2003). Zo representeren we bijvoorbeeld redelijk specifieke visuele eigenschappen van objecten en acties, zoals de vorm of oriëntatie van objecten en de richting waarin acties plaatsvinden (bv. Richardson, Spivey, Barsalou, & McRae, 2003; Stanfield & Zwaan, 2001). Maar kunnen sommige representaties nu ook inderdaad meer levendig of gedetailleerd zijn dan andere? En als dat zo is, kan taal dan voor deze verschillen zorgen? In **Hoofdstuk 4** onderzoeken we of taal de levendigheid van visuele representaties kan beïnvloeden. Proefpersonen bestudeerden een foto, waarna ze een kort verhaaltje lazen over het object op de foto. In dit verhaal waren de aanwezigheid en de relevantie van het object gemanipuleerd. In het voorbeeld van Matilda en de krijtjes lazen ze bijvoorbeeld een verhaal waarin de krijtjes wel of niet belangrijk waren voor Matilda's doel om mevrouw Bulstronk angst aan te jagen (relevantie), en waarin er wel of geen krijtjes in het klaslokaal lagen (aanwezigheid). Eerdere studies hebben aangetoond dat aanwezige en relevante voorwerpen meer actief zijn in ons geheugen dan afwezige en irrelevante voorwerpen (bv. MacDonald & Just, 1989; Suh & Trabasso, 1993; Zwaan & Radvansky, 1998). Echter, wij kozen deze twee factoren omdat beschreven aan- of afwezigheid de visuele aanblik van een situatie zou veranderen (iets is er wel of niet), terwijl beschreven relevantie dit niet doet (of iets belangrijk is of niet verandert niet hoe een situatie er uit ziet). Na het lezen van het verhaal kozen de proefpersonen uit twee foto's de foto die het meest leek op de eerder gepresenteerde foto. De foto's waren hetzelfde met als enige verschil dat de een transparanter (minder levendig) was dan de foto die ze eerder zagen en de ander minder transparant (levendiger) was. De resultaten van twee experimenten lieten zien dat mensen bij het lezen over een afwezig object vaker het transparantere plaatje kozen dan bij het lezen over een aanwezig object. Met andere woorden, lezen dat er geen krijtjes waren in Matilda's klaslokaal lijkt te zorgen voor een minder levendige visuele representatie van krijtjes dan lezen dat er wel krijtjes waren. Dit suggereert dat objecten die afwezig zijn in een beschreven situatie minder zichtbaar zijn in de bijpassende representatie. De relevantie van een object had geen effect op het gekozen plaatje. Het lezen dat Matilda de krijtjes niet of juist wel nodig had om mevrouw Bulstronk angst aan te jagen beïnvloedde de gekozen transparantie dus niet. Deze bevindingen laten zien dat de visuele levendigheid van concepten in het geheugen kan variëren, afhankelijk van de visuele situatie die beschreven is in taal.

In **Hoofdstuk 5** stellen we een gerelateerde vraag, namelijk hoe specifiek de wisselwerking is tussen visuele en linguïstische informatie. In taal bestaan er verschillende niveaus van abstractie, bijvoorbeeld een hoog niveau (fruit), een basisniveau (appel), en een laag niveau (Granny Smith)(Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Wij bestudeerden of deze hiërarchie in taal zijn sporen achterlaat in de manier waarop visuele informatie gerepresenteerd wordt. Eerder onderzoek toonde aan dat mensen algemene informatie over categorieën (zoals fruit) opslaan in termen van specifieke voorbeelden (bv. appel) wanneer een zin dat toelaat (Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976). Dit onderzoek werd gedaan met enkel talige stimuli, namelijk met zinnen en woorden. Wij wilden bestuderen of dit effect van *specificiteit* terug te vinden is in een wisselwerking tussen taal en visuele informatie, namelijk met zinnen en tekeningen. In onze experimenten kregen de proefpersonen bijvoorbeeld een van de volgende twee zinnen te lezen:

The fruit forms an ingredient for cider. (1)
The fruit forms an ingredient for exotic salads. (2)

Beide zinnen noemen de categorie *fruit*, maar zin 1 stuurt naar het specifieke voorbeeld *appel* en zin 2 niet. Ons eerste experiment repliceert eerder onderzoek met zinnen en woorden, door aan te tonen dat mensen zich zin 1 beter konden herinneren dan zin 2 na het zien van het woord *appel*, wat suggereert dat ze specifiek een appel representeerden bij het lezen van zin 1 maar niet zin 2. In ons tweede experiment vonden we soortgelijke resultaten met gedetailleerde tekeningen. Bij het zien van een gedetailleerde tekening van een appel waren mensen beter in het zich herinneren van zin 1 dan zin 2 (en dit effect was groter dan bij meer schematische tekeningen van een appel). Dit laat zien dat er een wisselwerking is tussen het detail in visuele informatie en het detail waarin we ons taal herinneren. Het derde experiment onderzocht of deze wisselwerking plaatsvond tijden het opslaan (lezen) van de zin of tijdens het ophalen (zich herinneren) van de zin. Hier vonden we dat mensen sneller en accurater een gedetailleerde tekening van een appel herkenden na het lezen van zin 1 dan zin 2, wat niet het geval was bij een meer schematische tekening. Dit suggereert dat er al tijdens het lezen van zin 1 gedetailleerdere visuele informatie werd gerepresenteerd dan tijdens het lezen van zin 2, wat erop duidt dat specifiek geactiveerde voorbeelden zoals *appel* gepaard gaan met meer visueel detail dan het algemeen activeren van de categorie *fruit*. Hiermee hebben we aangetoond dat het detail van visuele representaties wordt beïnvloedt door de specificiteit die geïmpliceerd wordt in taal.

Samenvattend laten deze twee hoofdstukken zien dat (a) een beschreven situatie de levendigheid van visuele informatie in het geheugen kan beïnvloeden en (b) de specificiteit waarmee een situatie beschreven wordt het detail van een visuele representatie kan beïnvloeden. Deze studies dragen op verschillende manieren bij aan de literatuur. Allereerst combineren ze oudere inzichten uit het taalonderzoek (over bijvoorbeeld het representeren van aanwezigheid, relevantie, categorieën en voorbeelden) met nieuwe inzichten dat representaties verankerd moeten zijn in perceptie. De bevindingen kunnen inzichten uit het talige domain generaliseren naar het visuele domein. Ten tweede laten de resultaten zien dat het ophalen van herinneringen aan visuele en talige ervaringen die eerder in het experiment plaatsvonden, beïnvloedt wordt door het verwerken van informatie uit een ander domein (talig en visueel, respectievelijk). Bijvoorbeeld, verschillen in linguïstische informatie (verhaaltjes) zorgden voor verschillen in hoe de eerder ervaren visuele informatie werd opgehaald. Met andere woorden, verschillen in talige informatie zorgden er voor dat bepaalde visuele informatie meer of minder toegankelijk werd in het geheugen. Ten derde hebben we deze toegankelijkheid gemeten op een manier die inzicht gaf in de levendigheid en het detail van visueel gerepresenteerde informatie. De resultaten lieten zien dat beter toegankelijke visuele informatie ook meer levendig en gedetailleerd gerepresenteerd wordt. Hiermee toonden we aan dat de kwaliteit van visuele representaties afhangt van taalgebruik.

Conclusie

Het lezen van een goed boek of het luisteren naar een mooi verhaal kan je wegvoeren uit het hier en nu en kan er voor zorgen dat je een andere wereld voor je ziet (oftewel 'verbeeldt'). Dit fenomeen duidt er al op dat visuele informatie een rol speelt bij het verwerken van taal. In dit proefschrift hebben we het mechanisme onderzocht dat ten grondslag ligt aan deze ervaring, door te bestuderen hoe talige en visuele informatie elkaar beïnvloeden tijdens perceptie (waarneming) en in het geheugen. We vonden dat tweetaligen hun aandacht richten op visuele informatie die te maken heeft met de taal die ze op dat moment niet gebruiken. Verder lieten we zien dat het detail en de levendigheid waarmee visuele informatie wordt gerepresenteerd afhangt van de situatie die beschreven wordt door taal. Voortbouwend op de bestaande literatuur tonen deze bevindingen aan dat mentale representaties (situatiemodellen) visuele informatie bevatten. Deze visuele informatie werd al opgeslagen tijdens eerdere ervaringen in de omgeving en wordt opnieuw geactiveerd bij het verwerken van taal. Dit impliceert dat taalgebruikers de wondere wereld die wordt beschreven in een verhaal tot op zekere hoogte daadwerkelijk waarnemen, door het activeren (ophalen) van eerdere waarnemingen uit het geheugen en deze om te smelten tot een representatie van de beschreven situatie terwijl het verhaal zich ontvouwt.

Curriculum Vitae

Lisa Vandeberg was born in Venlo on August 31st, 1982. She completed secondary education in 2000 at College den Hulster in Venlo. After entering Linguistics at the Radboud University Nijmegen in 2000, she switched to Psychology in 2001. During this study, she spent four months in Montpellier, France, to perform her Master's research on multilingualism at l'Université Paul Valéry. She obtained her Bachelor's and Master's degree in cognitive psychology in 2006 and 2007 respectively (the latter cum laude). In 2008 she started as a Ph.D. student at the Erasmus University Rotterdam where she studied language from a situated and grounded perspective, of which the present thesis is the result. In addition to performing research, she supervised several Bachelor and Master theses and was involved in teaching a number of practical and theoretical courses, mainly in methodology and statistics. In September 2012 she will start working as a PostDoc at the Amsterdam School of Communication Research at the University of Amsterdam, where she will study how advertising affects implicit memory and behavior.

Papers

Vandeberg, L., Eerland, A., & Zwaan, R.A. (2012). Out of sight, out of mind: Language affects perceptual vividness in memory. *PLOS ONE*, *7*(4): e36154. doi:10.1371/journal.pone.0036154. (Chapter 4)

Vandeberg, L., Guadalupe, T., & Zwaan, R.A. (2011). How verbs can activate things: Cross-language activation across word classes. *Acta Psychologica, 138,* 68-73. doi:10.1016/j.actpsy.2011.05.007. (Chapter 2)

Vandeberg, L., Bouwmeester, S., Bocanegra, B.R., & Zwaan, R.A. (under revision). Detecting cognitive interactions through eye movement transitions. *Cognitive Psychology*. (Chapter 3)

Vandeberg, L., & Zwaan, R.A. (submitted). Seeing the apple in the fruit: Implied linguistic specificity affects the detail of visual representations. *PLOS ONE*. (Chapter 5)

Zwaan, R.A., & Vandeberg, L. (in preparation). Inhibition effects in grounded cognition.

* Additional: Vandeberg, L. (2009). Dutch translation of the Automated Reading Span Task (Unsworth, Heitz, Schrock, & Engle; 2005):
http://psychology.gatech.edu/renglelab/translations.html