Acta Psychologica 137 (2011) 151–171

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Using the visual world paradigm to study language processing: A review and critical evaluation

Falk Huettig [a,b,*], Joost Rommers [a,c], Antje S. Meyer [a,d]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, The Netherlands
[c] Radboud University, Nijmegen, The Netherlands
[d] University of Birmingham, UK

ARTICLE INFO

Article history:
Received 21 March 2010
Received in revised form 18 November 2010
Accepted 19 November 2010
Available online 1 February 2011

PsycINFO classification:
2300 Human Experimental Psychology
2326 Auditory & Speech Perception
2340 Cognitive Processes
2346 Attention
2720 Linguistics & Language & Speech

Keywords:
Attention
Language
Vision
Eye movements
Visual world paradigm

ABSTRACT

We describe the key features of the visual world paradigm and review the main research areas where it has been used. In our discussion we highlight that the paradigm provides information about the way language users integrate linguistic information with information derived from the visual environment. Therefore the paradigm is well suited to study one of the key issues of current cognitive psychology, namely the interplay between linguistic and visual information processing. However, conclusions about linguistic processing (e.g., about activation, competition, and timing of access of linguistic representations) in the absence of relevant visual information must be drawn with caution.

Contents

* Corresponding author. Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Tel.: +31 24 3521374.
E-mail address: falk.huettig@mpi.nl (F. Huettig).

0001-6918/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.11.003

1. Introduction

In 1974, Cooper asked participants to listen to short narratives while looking at displays showing common objects, some of which were referred to in the spoken text. The participants were informed that their pupil size was recorded and that they could look anywhere they wanted. In spite of these instructions, Cooper found that the listeners' gaze was drawn to objects that were mentioned or were in some way associated with the text. For instance, the listeners were more likely to look at a picture of a dog when hearing "my scatter-brained dog Scotty…" than during other passages of text, and their gaze was attracted to the picture of a camera when they heard "During a photographic safari…". Cooper also found that the listeners' eye movements were closely time-locked to the text, with more than 90% of the fixations to the critical objects being triggered either while the corresponding word was spoken or within 200 ms after word offset. Cooper felt that he had found a "practical new research tool for the real-time investigation of perceptual and cognitive processes and, in particular, for the detailed study of speech perception, memory, and language processing" (p. 84). However, Cooper's study was largely ignored by the psycholinguistic community for more than twenty years (being cited only eight times until 1996; 105 times until 2010). It was only after Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) published a *Science* paper using a similar methodology (see also Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995) that psycholinguists began to exploit the systematic relationship between eye movements and speech processing on a larger scale.

The paradigm pioneered by Cooper and by Tanenhaus and colleagues is now known as the visual world paradigm (Allopenna, Magnuson, & Tanenhaus, 1998) and has had a transformative impact on the field of psycholinguistics. One may ask why Cooper's study failed to get noticed, whereas 20 years later, the Tanenhaus et al. paper had such an enormous impact. In part this may be due to the fact that until the mid-nineties eye tracking was a rather cumbersome technique to use. In addition, the rise of the visual world paradigm reflects the theoretical development in psycholinguistics. Since the early eighties – when Fodor (1983) developed the notion of the modularity of mind – a key concern in psycholinguistics has been to determine how linguistic and non-linguistic processes jointly determine the listener's or reader's understanding of sentences. Tanenhaus and collaborators were the first to show that the visual world paradigm is a powerful tool to investigate this issue.

Soon after the first comprehension studies eye-tracking also began to be used to study language production (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998). Although production studies tend not to be labeled 'visual world research', there are obvious similarities to comprehension studies using the visual world paradigm, and we will therefore review both comprehension and production studies.

We will first outline the key features of the visual world paradigm (Section 2) and then review the main research areas where it has been used. In Section 3, we describe and evaluate visual world studies at the sentence and discourse level and in Section 4 we review studies at the word level. In Section 5, we review three lines of production research that have used eye tracking. A central point in our conclusion (Section 6) and throughout the review is how the presence of relevant visual information affects the processing of spoken language and, related to this, which conclusions can be drawn about spoken language processing in other situations.

2. Key properties of the visual world paradigm

In this section, we first (in Section 2.1.) review the typical properties of the paradigm as used in studies of language comprehension and then, in Section 2.2 turn to the properties of production studies.

2.1. Key properties of comprehension visual world studies

2.1.1. Display types and spoken utterances

The basic set-up in a comprehension visual world experiment is simple: On each trial the participants hear an utterance while looking at an experimental display. The participants' eye movements are recorded for later analyses. In one popular version of the paradigm the visual input consists of line drawings of semi-realistic scenes shown on a computer screen and sentences that describe or comment upon the scenes (e.g., "The boy will eat the cake", Altmann & Kamide, 1999; see Fig. 1i). Typically, the display includes objects mentioned in the utterance (i.e., a boy and a cake for the previous example) and distractor objects that are not mentioned. In another version the displays are sets of objects, either laid out on a workspace (e.g., Tanenhaus et al., 1995) or shown as line drawings on a computer screen (e.g., Allopenna et al., 1998; see Fig. 1ii). The use of semi-realistic scenes allows researchers to assess, among other things, how the listeners' perception of the scene and/or their world knowledge about scenes and events affect their understanding of the spoken utterances (for further discussion see Henderson & Ferreira, 2004). When arrays of objects are used, the impact of such knowledge is minimized, which renders arrays well suited for studying the activation of conceptual and lexical knowledge associated with individual words.

In some studies, a visual display was shown first, and a spoken sentence followed, while a blank screen was shown. Such a set-up is useful to investigate effects of short-term visual memory on language-mediated eye gaze. Blank-screen studies have shown that people tend to re-fixate the regions on the blank screen that were previously occupied by relevant objects suggesting that language-mediated eye movements are not contingent upon a visual item being co-present during that expression (see Altmann, 2004; Altmann & Kamide, 2007; Ferreira, Apel, &

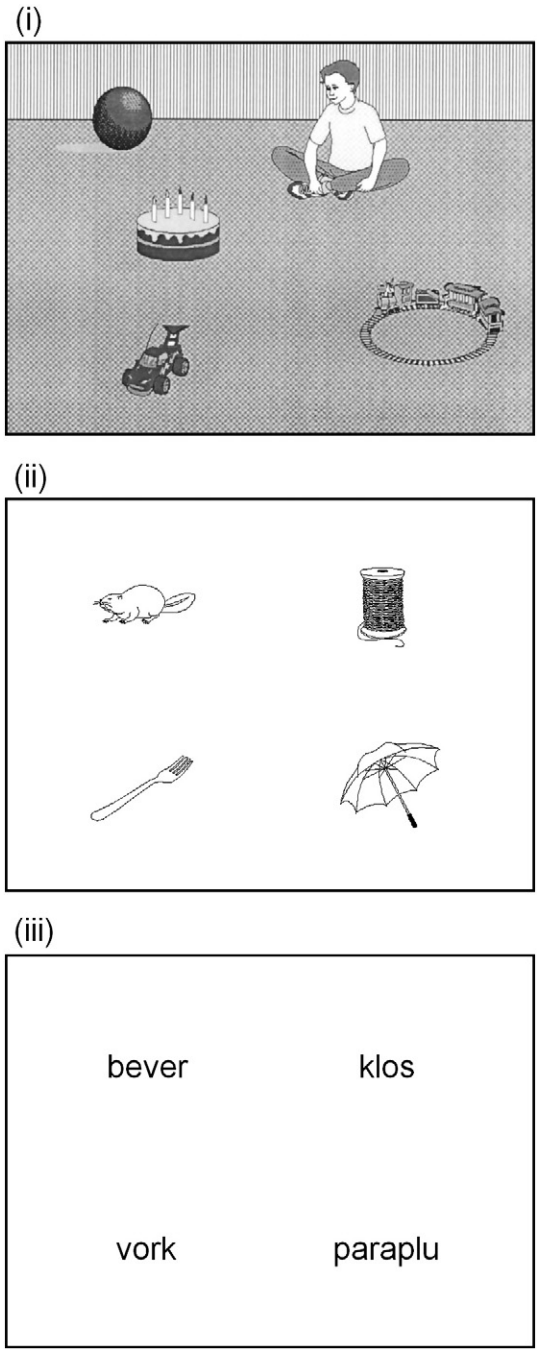

Fig. 1. Typical visual displays in the visual world paradigm. Example (i) is from Altmann and Kamide (1999), examples (ii) and (iii) are from Huettig and McQueen (2007). (i) Is an example of a semi-realistic scene (participants either heard "The boy will eat the cake" or "The boy will move the cake"). (ii) is an example of a four object display and (iii) is an example of a printed word display. In Examples (ii) and (iii) participants heard the Dutch sentences "Uiteindelijk keek ze naar de beker die voor haar stond" (Eventually she looked at the beaker that was in front of her).

Henderson, 2008; Knoeferle & Crocker, 2007; Richardson, Altmann, Spivey, & Hoover, 2009; Richardson & Spivey, 2000; Spivey & Geng, 2001).

Recently a printed word version of the visual world paradigm has been developed (Huettig & McQueen, 2007; McQueen & Viebahn, 2007; see Fig. 1iii). The only difference to the standard version is that the pictures are replaced by printed words. An important advantage of the printed word version is that the visual stimuli need not represent concrete objects but can be any words. Huettig and McQueen (2007) suggested that the printed word version may be more sensitive to phonological manipulations than the traditional version using pictures. Results obtained by Weber, Melinger, and Lara Tapia (2007) support this view. Salverda and Tanenhaus (2010) found that the printed word variant is a useful tool to investigate orthographic processing during speech perception. By contrast, it appears to be less sensitive to investigate the processing of semantic and (conceptual) visual-form representations than the standard version (Huettig & McQueen, 2008).

Many visual world studies have examined whether items that are phonologically, semantically or visually related to a target attract attention. To this end it is sometimes useful to include target-absent displays that do not feature the object mentioned in the accompanying sentence. For instance, Huettig and Altmann (2005) showed that this greatly increased the likelihood of observing competition effects (cf. Fig. 2 in Huettig & Altmann, 2005).

2.1.2. Stimulus timing

Researchers are often particularly interested in what happens during or immediately after the presentation of a critical word in the speech stream. Since this word typically appears in a carrier sentence ("Click on the <critical word> or "The boy will <critical word>...") participants have a few seconds to familiarize themselves with the objects prior to its onset. Usually the presentation of the visual display begins simultaneously with or shortly (e.g., 1 s) before the onset of the spoken utterance and stays in view until the end of the utterance. The amount of preview given is important. Huettig and McQueen (2007) showed that the likelihood of fixating particular objects depended on the time participants were given to retrieve relevant representations about the objects.

2.1.3. The task

The spoken utterances can be instructions to the participants ('direct action' tasks, e.g., "Pick up the candy", Allopenna et al., 1998) or mere descriptions or comments on the display ('look and listen' tasks, e.g., Altmann & Kamide, 1999; Huettig & Altmann, 2005). In the latter case, the participants are often asked to look at the screen and to listen carefully to the sentences. The choice of task depends on the experimental questions (see Huettig, Olivers, & Hartsuiker, this issue; Salverda, Brown, & Tanenhaus, this issue, for further discussion). 'Look and listen' tasks, for instance, allow researchers to evaluate whether particular (e.g., competition) effects are a more general feature of language–vision interactions or whether they are limited to certain specific task demands.

An advantage of the visual world paradigm compared to other psycholinguistic paradigms such as word spotting, lexical decision, and grammaticality judgments (see special issue of *Language and Cognitive Processes*, Grosjean & Frauenfelder, 1996) is that the listeners do not have to perform any meta-linguistic judgments, which might be difficult to elicit from some groups of listeners (e.g., young children) and might affect the way the speech is processed. The visual world paradigm solely relies on the listeners' tendency to look at relevant parts of the display as they are mentioned.

One may ask why listeners look at the objects that are mentioned or implied. It is not surprising that participants look at the target objects when they are asked to touch or move them; such tasks are accomplished most easily by fixating upon the object (see also the research on eye movements during everyday tasks such as making a cup of tea, e.g., Land, Mennie, & Rusted, 1999; Irwin, 2004). However, listeners also look at the target objects when no overt action is required. Most likely they aim to relate the spoken utterance to the visual input (e.g., Altmann & Kamide, 2007). Perhaps they do so because this kind of mapping is expected and beneficial in many everyday contexts, for instance when we study a text book including diagrams and figures, when we listen to an instructor explaining and demonstrating a new skill (e.g. planting tulip bulbs or roasting a duck), or when we watch the weather forecast on TV. Very often the spoken and visual media provide complementary information, and it is useful to process them together. Relating the visual and spoken information to each other is achieved most easily by directing one's visual attention – and eye gaze – to the relevant objects. This is because

attending to an object facilitates not only the recognition of the object, but also the activation of any associated information, including, for instance, the object's name (e.g., Malpass & Meyer, 2010).

Whether the integration of the visual and spoken input is a deliberate or rather automatic process remains to be seen. Current influential accounts regard automaticity as a continuum (e.g., Logan, 1985) and favor the view that automaticity can be best diagnosed by looking at the presence or absence of compositional features such as intentionality, controllability, goal dependence, efficiency, and whether the process in question is purely stimulus driven, unconscious, and fast (see Moors & De Houwer, 2006, for a comprehensive review).

Altmann and Kamide (2007) have proposed that an increase in the activation of a mental representation of an object and its location (e.g., by linguistic input) results in the increased likelihood of a saccadic eye movement towards this location. Huettig et al. (this issue, cf. Knoeferle & Crocker, 2006; Spivey, Richardson, & Fitneva, 2004) proposed that *working memory* serves as the nexus where long-term visual and linguistic representations (i.e. types) are bound to specific locations (i.e. tokens or indices). This is similar to what Altmann and Kamide (2007) describe as episodic traces of the experience of an object ("including its location and the conceptual representations associated with that experience", p. 512). The main difference is that Huettig et al. assume that working memory is a necessary condition for this experience (see Huettig et al., this issue, for further discussion) whereas Altmann and colleagues (Altmann & Mirković, 2009) argue that working memory is not required as an external mechanism "because attention is instantiated within the same representational substrate as linguistic and nonlinguistic information (see also Cohen, Aston-Jones, & Gilzenrat, 2004)" and that therefore "different states of this representational substrate represent the attentional modulation that drives eye movements" (p. 593).

In sum, the listeners' eye movements during a trial of a visual world experiment reflect the direction of their visual attention, which depends both on the visual and auditory input. In other words, although the visual world paradigm is usually used to study speech comprehension, the listeners' eye movements do not reflect exclusively their linguistic processing but depend on their visual processing as well. Therefore, it is, for instance, conceivable that a target word is recognized earlier in a visual world experiment, where its phonological representation is activated not only by the spoken utterance but also by the pictorial representation of the referent (e.g., Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarette & Costa, 2005), than in the absence of a visual representation of the referent object. Although this may seem entirely obvious, the ways in which visual and auditory information jointly determine attention and gaze have, in our view, not received sufficient attention in the literature. We will return to this issue throughout this review.

2.1.4. Participants

In most visual world studies, participants have been tested individually, but there are also some studies which tested interlocutors in dialogue. These studies have used confederates instructing the participants (e.g., Hanna, Tanenhaus, & Trueswell, 2003; Keysar, Barr, Balin, & Brauner, 2000), naïve participants who were assigned a listener or speaker role (e.g., Snedeker & Trueswell, 2003), or participants in collaborative games (e.g., Brown-Schmidt, Gunlogson, & Tanenhaus, 2008, see Section 3.2). Participants in visual world studies have mostly been undergraduate students, but the paradigm has also been used in patient groups (Thompson & Choy, 2009; Walsh, Dickey, Choy, & Thompson, 2007; Yee, Blumstein, & Sedivy, 2008) and in adolescents (Brock, Norbury, Einav, & Nation, 2008; McMurray, Samelson, Lee, & Tomblin, 2010). It is, of course, closely related to the preferential looking paradigm often used in developmental studies (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987, see also Arias-Trejo & Plunkett, 2010; Swingley & Fernald, 2002; Johnson & Huettig, 2011; Nation, Marshall, & Altmann, 2003; Trueswell, Sekerina, Hill, & Logrip, 1999).

2.1.5. Data analysis

The data analyses in visual world studies focus on the question of how likely the participants are to look at specific regions of interest at different times during a trial. The selection of the regions of interest and of the time windows depends on the research question. The regions of interest might, for instance, be the drawings of a target object mentioned in the utterance and of a distractor object with a similar sounding name, and the time windows might be 100-ms-episodes starting from the onset of the name of the target object. The most common dependent variables are fixation proportions on the interest areas during each time window or counts of saccades towards the regions of interest initiated during each time window (e.g., Altmann, 2004).

Typical questions for the statistical analyses are whether two regions of interest differ in their likelihoods of being inspected during each of a set of consecutive time windows (e.g., Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Huettig & McQueen, 2007), or whether a region of interest (e.g., a cake) is looked at earlier in an experimental condition (e.g., when it is implied by a verb, as in "The boy will eat …") than in a control condition ("The boy will move …", e.g., Altmann & Kamide, 1999). In most studies the proportions of fixations or saccades are compared in *t*-tests or analyses of variance (often after suitable transformations). These analyses have proven to be quite robust (as evident from the numerous successful replications). Note however that visual world data violate some of the underlying assumptions of ANOVAs (see Barr, 2008a, 2008b). The assumption of the independence of observations is violated if time window is included as a factor because the same fixations may contribute to multiple time windows. The normal distribution assumption and the continuous variable assumption may be violated if proportions (i.e. categorical data) are the dependent variable. Alternative statistical approaches such as multilevel logistic regression (Barr, 2008a, 2008b), log-linear analysis (Huettig & Altmann, 2005; for discussion see Howell, 2002; Scheepers, 2003), and growth curve analyses (Mirman, Dixon, & Magnuson, 2008) are increasingly being applied to visual world data (see also the special issue of the *Journal of Memory and Language*, Volume 59, 2008).

A more serious statistical issue concerns baseline effects. For instance, a particular object (e.g., a competitor to the target) may be more likely to be fixated *before* the critical information from the spoken utterance is accessed. Such biases must be corrected in the statistical analysis (see Barr, Gann, & Pierce, this issue).

When interpreting the results of visual world studies, it is important to keep in mind that fixations and saccades are relatively discrete events. Thus data from a single trial cannot provide information about the continuous processing of the speech signal, e.g., the gradual activation of word candidates or the gradual deactivation of competitors. However, by averaging across trials and participants, it can be computed how likely listeners are, on average, at a given moment in time, to look at each of the areas of interest. Based on these data, inferences about the time course of the underlying cognitive processes can be drawn.[1] Increasingly, the results of visual world experiments are used not only to inform descriptive models of language processing, but also to develop and test computational models (Allopenna et al., 1998; Mayberry, Crocker, & Knoeferle, 2009; Mirman & Magnuson, 2009; McMurray et al., 2010; Roy & Mukherjee, 2005; see also Stephen, Mirman, Magnuson, & Dixon, 2009).

2.2. Key properties of production studies

In production studies, participants see sets of objects (e.g., Griffin, 2001; Meyer et al., 1998) or cartoons of events or actions (e.g.,

[1] Note that Spivey, Grosjean and Knoblich (2005), see also Farmer, Anderson, and Spivey (2007) recently proposed an alternative paradigm, where trajectories of a computer mouse are recorded while participants are carrying out instructions such as "click on the candle". They argued that trajectory information yielded a more direct way of measuring the continuous processing of auditory information (but see Van der Wel, Eder, Mitchel, Walsh, & Rosenbaum, 2005).

Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000; Griffin & Oppenheimer, 2006). No spoken input is presented, but instead the participants are asked to describe what they see. Sometimes, detailed instructions are given about the expected utterances; speakers might, for instance, be asked to name the objects in a specific order (e.g. left to right) in bare noun phrases; Meyer et al., 1998), and sometimes speakers are simply asked to describe what they see (Griffin & Bock, 2000). As in the comprehension studies, the participants' eye gaze is monitored. Researchers typically determine which objects are inspected, in which order they are inspected, and when they are inspected relative to the participants' speech output. As will be shown in more detail later, this provides information about the ways speakers coordinate the generation of utterance plans with the overt articulation. In addition, researchers often determine how long each object is gazed at. After a review of comprehension studies, we will show later (Section 5) that the duration of a speaker's gaze to an object is a good indicator of the time they need to identify the object and generate utterances about it.

3. Studies of language comprehension at the sentence and discourse level

3.1. Sentence processing

A key controversy in the study of language comprehension is how and when language users integrate different types of information. There are two contrasting theoretical views. According to structural (or two-stage) accounts, the listener's or reader's initial parsing of a sentence is based exclusively on syntactic information; other types of information (e.g., lexical and pragmatic information) exert their influence only at a later stage (Frazier, 1979, 1987). According to interactive theories (e.g., Tyler & Marslen-Wilson, 1977), non-syntactic information can immediately influence sentence processing. The currently most influential type of interactive models are constraint-based theories (MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). Constraint-based accounts assume that syntactic processing has similar properties to lexical processing (MacDonald et al., 1994; Trueswell et al., 1994) and is accomplished through the satisfaction of multiple constraints. These include lexical, structural, and discourse level constraints. Lexical constraints (e.g., argument structure preferences and frequency of co-occurrence of words within a phrase) however are assumed to provide stronger constraints than structural or discourse constraints (e.g., Trueswell & Tanenhaus, 1994). Much of the visual world research on sentence and discourse processing has been devoted to distinguish between structural and constraint-based theories.

3.1.1. Visual (and other) constraints on spoken sentence processing

Before the seminal study by Tanenhaus et al. (1995) most psycholinguistic investigations of sentence processing relied on the examination of reading times. Contextual manipulations were therefore limited and typically involved changing the preceding sentence or discourse (see Frazier, 1995; MacDonald et al., 1994; for review). There are, however, many other sources of information that language users could potentially exploit in their daily interactions, including, for instance, visual information constraining the intended meaning of utterances. To investigate when listeners make use of such information, Tanenhaus and colleagues presented participants with sentences such as "Put the apple on the towel in the box", where the first prepositional phrase ("on the towel" in the example) is temporarily ambiguous between denoting the destination of the apple or its current location. In the one-referent condition of the experiment participants saw just one apple on a towel, an empty towel, a box, and a pencil. In the two-referent condition there were two apples: one on a towel and one on a napkin. In this condition, a modifier was needed to inform the listener which of the two apples should be moved. According to structural accounts of syntactic ambiguity resolution (Frazier, 1987), the phrase "on the towel" should initially be interpreted as the destination of the apple, regardless of the visual context, because this is the structurally simplest syntactic analysis of the sentence. This should manifest itself in many early looks to the empty towel. However, Tanenhaus et al. (1995) found that there were significantly more early looks to the empty towel in the one-referent than in the two-referent condition. This is strong evidence that, contrary to two-stage accounts of sentence processing, listeners can use visual information immediately to disambiguate sentence structures.

Trueswell et al. (1999) conducted a similar study with adults and five-year-old children. For the adults, they replicated the context effect demonstrated by Tanenhaus et al., but the children were equally likely to look at the empty towel in the one-referent and in the two-referent condition. Thus, they failed to use the contextual information when they processed the sentences. Trueswell and Gleitman (2004) argued that young children can use visual context to guide syntactic choice, but that they need time to discover the usefulness of such information because contextual cues are not always present and tend to be less reliable than syntactic cues.

Snedeker and Trueswell (2004) provided further evidence that syntactic ambiguity resolution is accomplished through the satisfaction of multiple constraints. Participants heard ambiguous sentences such as "Tickle the pig with the fan" or "Choose the cow with the stick" in a one-referent and a two-referent condition. They found that the degree of preference for an instrument interpretation (rather than a modifier interpretation) of the ambiguous phrase, which was independently assessed, modulated eye gaze. These results support the view that both linguistic constraints and visual context can determine the initial syntactic analysis of sentences. Moreover, they show that the influence of particular cues may change over the course of development.

Chambers, Tanenhaus, and Magnuson (2004) provided a particular striking demonstration of early contextual influences. Participants heard instructions such as "Pour the egg in the bowl over the flour". Chambers et al. (2004) found that participants immediately interpreted "in the bowl" as the modifier when the scene contained two eggs in liquid form (one in a glass, the other in a bowl), as revealed by few looks to an empty bowl (the 'false goal' object); however when the scene contained one liquid egg (in a bowl) and one solid egg (in a glass), participants expected the phrase "in the bowl" to be the intended location for the single pourable egg (i.e. the flour); as revealed by increased looks to the 'false goal' object (i.e. the empty bowl). Thus affordances compatible with the action influenced the earliest moments of syntactic processing.

Arnold, Kaiser, and colleagues (e.g., Arnold, 2001; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Kaiser, Runner, Sussman, & Tanenhaus, 2009; Kaiser & Trueswell, 2008) used the visual world paradigm to study the interpretation of pronouns (e.g., she/he), demonstratives (e.g., this and that), and reflexives (e.g., herself/himself). Consistent with interactive theories, these studies provided evidence that reference resolution is sensitive to multiple constraints (e.g., information structure, syntactic role, and word order), whose impact differs across anaphoric forms. For instance, Kaiser et al. (2009) found that the resolution of pronouns was influenced more by semantic (and less by syntactic) information than the interpretation of reflexives. These studies therefore suggest a complex interaction of syntactic and semantic factors during reference resolution (see also Brown-Schmidt, Byron, & Tanenhaus, 2004, 2005).

In sum, the visual world studies reviewed in this section provide strong support for constraint-based accounts of sentence processing. To the best of our knowledge no visual world studies have reported evidence supporting structural accounts. Evidence for the immediate influence of multiple constraints in other situations (e.g., during reading) however appears to be more mixed (see Clifton & Staub, 2008; Clifton et al., 1994;

Clifton et al., 2003; Van Gompel & Pickering, 2001; Van Gompel, Pickering, Pearson, & Liversedge, 2005, for discussion; but see ERP evidence of Hagoort, Hald, Bastiaansen, & Petersson, 2004, and Sereno, Cameron, & O'Donnell, 2003). Mitchell (2004, p. 22), for instance has argued that there are "serious questions about the extent to which visual-world studies can throw light on core context-invariant sentence processing, and the degree to which the results merely reflect the kinds of processing that occur in the presence of particular object arrays" (p. 22). We will return to this question throughout the review.

3.1.2. Predictive understanding

A related line of research was initiated by Altmann and collaborators. Here listeners see semi-realistic scenes and hear sentences commenting about them. The key question is how and when linguistic information from the spoken sentences (e.g., "She will pick up the bottle and pour the wine carefully into the glass") is integrated with information retrieved from the visual environment (e.g., a scene depicting a woman, a table, a wine bottle, and an empty glass). This work has (i) demonstrated the importance of prediction during language processing and (ii) shown that language-mediated eye movements do not only reflect linguistic processing but also the constant updating of dynamically changing mental representations of the event that the scene *and* the spoken utterance refer to.

Altmann and Kamide (1999) presented participants with semi-realistic visual scenes depicting, for instance, a boy, a cake, and some toys (see Fig. 1i) while they heard sentences such as "The boy will move the cake" or "The boy will eat the cake". They found that eye movements to the cake (the only edible object in the scene) started significantly earlier in the "eat" condition than in the "move" condition. Altmann and Kamide (1999) interpreted this result as evidence that selectional information conveyed by a verb can be used to anticipate an upcoming theme. Kamide, Altmann, and Haywood (2003) explored whether only verb information can be used to predict what will be referred to next or whether the combination of verb information with the preceding grammatical subject can drive anticipatory eye movements. They found increased fixations to a motorbike when participants heard "The man will ride …" but increased fixations to a carousel when participants heard "The girl will ride …". Thus, information provided by the grammatical subject and by the verb can jointly constrain anticipatory eye movements (see also Kamide, Scheepers, & Altmann, 2003, for evidence that case-marking information can be used for prediction). Altmann and Kamide (2007) showed that tense information is used to interpret which referent is being referred to. Their participants tended to look at an *empty* wine glass when hearing the past-tense sentence "The man has drunk …" but at a full glass of beer when hearing "The man will drink …".

Recently, Altmann and Kamide (2009) showed that listeners' anticipatory and concurrent eye movements depend not so much on the properties of the spoken input and the visual scene but rather on the mental representation the listeners construct while processing the stimuli over time (cf. Kaiser & Trueswell, 2004). Participants heard sentences such as "The woman will put the glass on the table" or "The woman is too lazy to put the glass on the table" while viewing a scene featuring a woman, a wine bottle and a wine glass on the floor and an empty table. They then heard "The woman will pick up the bottle and pour the wine carefully into the glass". In Experiment 1 the display remained on the screen, in Experiment 2 it was replaced by an empty screen before sentence onset. On hearing "pour" the participants of both experiments often looked at the location of the glass implied by the utterance (i.e. either on the floor or on the table). These studies therefore suggest that language-mediated eye movements reflect dynamically updated event representations (see also Knoeferle & Crocker, 2006, 2007, and Knoeferle et al., 2005, for related work contrasting the role of visual information and world knowledge of events in interpreting sentences; and Mishra & Singh, 2010, for work on fictive motion understanding).

In sum, listeners use a wealth of linguistic as well as visual information to disambiguate different sentence structures and to predict the upcoming linguistic input. Language-mediated eye movements reflect neither only linguistic processing nor only processing of the visual scene, but they reflect continuously updated mental representations based on information derived from both the linguistic and the visual input.

The visual world paradigm has proven to be a very useful tool to investigate cognitive processing: The participants' ongoing syntactic analyses, predictions, and event representations can be inferred from the direction of their eye gaze. One important issue, however, is how results in the visual world paradigm relate to syntactic analyses, predictions, and event representations in the absence of visual input. As Kamide, Altmann et al. (2003), explain "What we do not know, then, is whether for a fragment such as "The man will ride …", the processor activates representations at ride that correspond to the range of rideable objects that a man might ride even in the absence of a concurrent visual context portraying one or more rideable things. Thus, the paradigm does not allow us to determine whether it is the linguistic structure that triggers a predictive process, because the syntax determines that something is about to be referred to, or whether it is the visual context that suggests to the processor that something might plausibly enter into a thematic relationship with the man, mediated by the act of riding. We do not, needless to say, have data that empirically address this distinction one way or the other." (p. 151; see also Kamide, 2008, and Kamide, Scheepers et al., 2003, for discussion of this issue). What these studies show is what listeners *can* do, not what they actually do in other contexts.

3.2. Pragmatics and dialogue

Another line of research has applied the visual world paradigm to issues of pragmatics: the study of those aspects of interpretation that go beyond semantics and require inferences about the context and the speaker's goals. Three main pragmatic themes have been addressed: the use of context for situation-specific interpretations, the time-course of pragmatic inferences, and the use of common ground in dialogue situations.

3.2.1. Situation-specific interpretation

The visual world paradigm, by its nature, provides participants with a visual context with respect to which the linguistic input can be interpreted. Chambers et al. (2002) showed that this context is immediately used for situation-specific interpretations of prepositions: *inside* in "Put the cube inside the can" caused participants to restrict their visual attention immediately not just to containers, but more specifically to containers that were large enough to hold the specific cube present in the visual context. Episodic memory of the situation can also guide comprehension. Chambers and San Juan (2008) demonstrated that when participants had moved an object to a different location on a grid, the verb *return* (in, e.g., "Return the boot to area three") led to anticipatory saccades to this previously displaced object. This confirms that listeners' predictions are not only based on stable properties of objects (such as edibility; Altmann & Kamide, 1999), but can also be based on episodic knowledge about the situation at hand. Communicative relevance constrained this anticipatory effect: In a follow-up experiment employing a referential communication task, it was only found when the listener believed that the object's displacement was known by the speaker giving the instructions. Thus, the visual world paradigm is well suited for investigations of situational influences on language processing.

3.2.2. Pragmatic inferencing

Several studies have used the paradigm to address the time-course of pragmatic inferences, focusing on classical Gricean maxims (Grice,

1975) such as the maxim of quantity: provide as much information as necessary and not more. This maxim dictates that over-descriptions should be avoided. Indeed, Engelhardt, Bailey, and Ferreira (2006) found that over-descriptions led to eye movements suggesting confusion on the listener's side. On the other hand, speakers should not provide too little information: Reference to a particular glass in the context of other glasses should include information that uniquely distinguishes the glass in question from the other glasses (e.g., an adjective: "the tall glass"). Sedivy, Tanenhaus, Chambers, and Carlson (1999) showed that when such an adjective is provided by the speaker, the pragmatic cue is quickly taken into account. Upon hearing "Pick up the tall glass", participants were faster to direct their eye gaze to the target object when a "contrast object" (a short glass) was also present in the display than when no contrast object was present in the display. They did so at or even before the onset of the noun, supporting incremental views of language processing in which pragmatic inferences are drawn rapidly. A recent study (Grodner & Sedivy, in press) showed that this contrast inference was only drawn when the speaker was reliable: In the case of an unreliable speaker who faked "language and social problems", participants were not aided by the contrast, suggesting that the processes tapped into in these studies were truly pragmatic inferences rather than the automatic application of stored knowledge of conventions.

Even though pragmatic inferences may be drawn quickly, at an earlier stage some degree of lexical-semantic processing is likely necessary (i.e. there may be a lag between lexical-semantic and pragmatic processing). In order to tap into this semantics–pragmatics interface, Huang and Snedeker (2009a) asked participants to listen to utterances such as "Point to the girl that has some of the socks" while viewing a display in which one girl had two of four socks and another girl had three of three soccer balls (a phonological onset overlap competitor). The lexical semantics of "some" denote a quantity greater than one (i.e., *some-and-possibly-all*), but the word is usually interpreted with an 'upper boundary' (i.e., *some-and-not-all*) via a pragmatic inference: "some" does not mean *all* because if the speaker was in the position to say "all" he would not use "some". They found that the proportions of looks to the target were lower when "some" was presented than it was replaced with *all* or with numbers (*two* and *three*), which have exact semantics (Huang, Spelke, & Snedeker, 2005). This suggests that computing pragmatic inferences takes time. In fact, instead of calculating the pragmatic inference participants waited until the acoustic information disambiguated the target (soccer balls or socks; see Panizza, Chierchia, Huang & Snedeker, in press, for a similar result), but with more time between the introduction of the ambiguity and the disambiguation, the late inference was visible (Huang & Snedeker, in press). A view in which pragmatic inferencing is time-consuming and potentially effortful is also supported by developmental research showing that five-year-old children fail to draw certain pragmatic inferences, interpreting only the semantic content of quantifiers (Huang & Snedeker, 2009b). However, a recent study suggested that sufficient contextual support, including an instruction enhancing the salience of the item sets, can remove pragmatic inferencing delays (Grodner, Klein, Carbary, & Tanenhaus, 2010), opening the possibility that the increase in processing time for scalar adjectives observed in earlier studies arises not at the inferencing level but rather at the level of integrating the inference with the context.

3.2.3. Dialogue

Another body of research has applied the visual world paradigm to dialogue situations. Brown-Schmidt and Tanenhaus (2008) employed a referential communication task in which pairs of participants together arranged a set of objects, each on their own board. They hypothesized that the coordination that emerges during conversation and is known to facilitate comprehension (e.g., Schober & Clark, 1989) would constrain the referential domain to a small area of the board, and that this could be observed on-line. Indeed, after hearing a noun (e.g., *cloud*), looks to cohort competitors (e.g., *clown*) increased much less during unscripted conversation compared to during experimenter instructions (e.g., "look at the cloud"), reflecting that cohort competitors were outside the referential domain established in the conversation (through, e.g., task-relevance heuristics or proximity to the last mentioned object) as compared to experimenter instructions in which no such referential domain was specified.

The visual world paradigm has also been used to study how speakers and listeners use shared knowledge or common ground. The paradigm allows for objects to be placed in common ground, visible to both participants, or in privileged ground, visible to only one of the participants. This property has been exploited for investigating at which point in time common ground is used to resolve ambiguous references; in other words, do we immediately take the knowledge of our interlocutors into account, or is this a late, perhaps effortful process?

Some studies provided evidence for the addressees' immediate use of common ground. Hanna et al. (2003) presented participants with instructions such as "Now put the blue triangle on the red one". The display contained, apart from the target (a red triangle in common ground), another red triangle which was in the addressee's privileged ground (i.e. only the addressee could see it). Hanna, Tanenhaus and Trueswell observed that 400 ms after the onset of the adjective ("red") addressees were already more likely to look at the common ground target compared to the privileged ground competitor (see also Hanna & Tanenhaus, 2004; and for similar effects with five- and six-year old children, Nadig & Sedivy, 2002). Thus, addressees quickly take common ground into account, exhibiting a preference for looking at objects which the speaker can also see as opposed to objects the speaker cannot see and could thus not have referred to.

In contrast, other studies support the idea that addressees do not completely restrict understanding to common ground, considering and sometimes even selecting referents in their privileged ground. For instance, in a study by Keysar et al. (2000) addressees had to interpret a speaker's instructions to move objects in an array. On critical trials, three variants of an object were present in the array (e.g., a small, a medium-sized, and a big candle, along with an unrelated distractor), with one object (e.g., the smallest candle) being in the addressees privileged ground. When asked to move "the small candle", addressees looked longer and more often at the object in privileged ground when it was a competitor (e.g., the smallest candle) compared to when it was a distractor. Keysar et al. found that in one fifth of the trials listeners even moved the unintended object, demonstrating that they sometimes use an 'egocentric' heuristic that does not take common ground into account (see Keysar, Lin, & Barr, 2003, for a similar result; for a review, see Barr & Keysar, 2006).

However, looks at referents in privileged ground can also indicate that common ground is actually taken into account: when listeners are asked a question about an object that the partner cannot see, it is appropriate to inspect entities in privileged ground. Brown-Schmidt et al. (2008) instructed participants to re-arrange drawings of animals with different accessories in an array, with some slots in the array being occluded for one of the participants. When addressees were asked a question (e.g., "What's above the cow with the hat?"), they were more likely to direct their gaze to their own (privileged) slots than to shared slots in common ground. This happened quickly after the onset of the critical word (e.g., "cow"), suggesting that listeners immediately distinguished between common and privileged ground.

These studies illustrate how the visual world eye-tracking paradigm can be used to study high level processes that play an important role in natural conversations. However, with respect to the interlocutors' use of common ground, the results are somewhat inconsistent. Several reasons for this lack of convergence have been suggested in the literature. For instance, some studies indicating early use of common ground have used interactive paradigms with two

naïve participants (e.g., Brown-Schmidt et al., 2008), whereas studies showing an initially egocentric perspective have used a confederate director (Keysar et al., 2000) or pre-recorded utterances combined with a cover story aimed to convince participants that they were listening to a naïve participant who spoke to them through an intercom (Barr, 2008a, 2008b). It has been suggested that using a confederate may eliminate "many of the natural collaborative processes that occur in interactive conversation" (Tanenhaus & Brown-Schmidt, 2008, p. 1114) and that "effects of perspective are likely to be strongest in tasks where participants have joint goals, common ground is established collaboratively, and exchange of information is negotiated by both parties" (Brown-Schmidt et al., 2008, p. 1133; see also Barr, 2008a, 2008b, for discussion). Indeed, there is evidence that participants take their interlocutor's perspective into account more systematically in interactions with other naïve participants than with confederates (see Schober & Brennan, 2003, for discussion of Brown & Dell, 1987, and of Lockridge & Brennan, 2001). Brown-Schmidt (2009a) also showed that an on-line partner-specific facilitatory effect of 'entrained terms' (shared names developed during conversation) was only obtained during an interactive task and not in listening to recordings, again supporting the idea that interactivity is important in evaluating effects of common ground. However, interactivity cannot explain all of the discrepancies between studies since some studies demonstrating early effects of common ground used a confederate (Hanna & Tanenhaus, 2004; Hanna et al., 2003) or partly scripted utterances (Heller, Grodner, & Tanenhaus, 2008).

Another suggestion is that the type of ambiguity that was used might explain the differences. For instance, Hanna et al. (2003) used identical objects (e.g., a red triangle) in common and privileged ground. In this case, the linguistic input never disambiguates the referent and therefore knowledge of common ground must be consulted, perhaps encouraging listeners to abandon any egocentric strategy (as suggested by Barr, 2008a, 2008b, p. 20). In contrast, Keysar et al. (2000) used different items in common compared to privileged ground, with the item in privileged ground always being the best match to the linguistic input (e.g., when instructed to move the small candle, the smallest candle was in privileged ground, and the candles in common ground were medium sized (target) and large sized). As pointed out by Hanna et al. (2003, p. 45), this match could have made it harder for addressees to ignore the items in privileged ground.

Two studies (Barr, 2008a, 2008b; Heller et al., 2008; similar to Brown-Schmidt et al., 2008) attempted a more subtle manipulation: temporary ambiguities. Heller, Grodner and Tanenhaus (2008) (cf. Sedivy et al., 1999) instructed listeners to find the referent of an expression with a scalar adjective (e.g., "Pick up the big duck" while a big and a small duck were present in common ground along with a 'competitor contrast' (e.g., a big and a small box) in their privileged ground. Listeners anticipated the target (e.g., the big duck) while hearing "big" in spite of the presence of the competitor contrast, indicating that they used common ground from the earliest moments. In contrast, Barr (2008a, 2008b) found that when listeners encountered a temporarily ambiguous referring expression (e.g., *bucket*), the amount of interference from a phonological competitor (e.g., *buckle*) was equivalent regardless of whether the competitor was in common or privileged ground. Before the onset of the referring expression, participants preferentially looked at objects in common ground; but apparently this information was not used during the subsequent integration processes (see Keysar et al., 2000, Experiment 2, for a similar observation). At present, it is unclear whether a similar dissociation between anticipation and integration can also explain Heller et al.'s results; perhaps a study combining the scalar adjective and cohort manipulations could elucidate this issue.

Finally, differences between the results obtained in the available studies may be due to differences in the samples tested. Brown-Schmidt (2009b) showed that individual differences in inhibitory control ability predicted the addressees' sensitivity to common ground: The better participants' inhibitory control, the less likely they were to fixate upon the competitors.

Overall, the discrepancies in the results of similar studies again underscores the importance for users of the visual world paradigm to take into account exactly how linguistic information, derived from different kinds of ambiguities, and visual information, provided by different combinations of objects in the visual display, along with the social aspects of the test situation might affect the results.

3.3. Prosody and disfluencies

Several studies have used the visual world paradigm to investigate how properties of the speech signal might be used to guide syntactic ambiguity resolution and the analysis of information structure. The first of these properties concerns the prosodic features of sentences, the other disfluencies.

Regarding the line of research into prosody, Dahan, Tanenhaus, and Chambers (2002) showed that listeners could use intonation to determine whether the speaker will introduce a new referent or refer again to a previously mentioned one (e.g., "Put the candle above the triangle. Now put the candle ..."). They found that listeners used pitch accent of vowels to direct eye gaze: When hearing words with accented vowels participants tended to look at a new (or a non-focused) object, whereas when they heard words with unaccented vowels they tended to look at the previously mentioned (or most salient) object. Weber, Braun and Crocker (2006) (see also Ito & Speer, 2008) showed that listeners shifted eye gaze earlier towards the picture of a referent belonging to a contrast pair ("red scissors", when there were red scissors and purple scissors in the display) when there was contrastive accent on the adjective than when the adjective was not accented (see also Watson, Tanenhaus, & Gunlogson, 2008).

Snedeker and Trueswell (2003) demonstrated that prosody influenced the listeners' interpretation of syntactically ambiguous phrases (e.g., "Tap the frog with the flower") even before the onset of the ambiguously attached preposition. Thus, similar to the use of syntactic and semantic information and world knowledge (see Section 3.1), it appears that listeners are able to use prosodic information to predict upcoming linguistic input. Overall, these studies show that listeners can combine prosodic cues and visual context to determine the intended referent, and that they do so immediately, while the spoken input is still unfolding over time.

Another line of research has investigated the role of speaker disfluencies. Arnold, Altmann, Fagnano and Tanenhaus (2004) examined whether listeners use the increased likelihood of speakers to be disfluent (saying, e.g., *thee uh candle* instead of *the candle*) while referring to new as compared to given information (Arnold, Wasow, Losongco, & Ginstrom, 2000) as a cue to the information structure of the utterance. They found that, compared to fluent speech, disfluencies led to more looks to discourse-new objects. In addition, disfluent instructions resulted in fewer looks to a cohort competitor (e.g., *camel*) of a discourse-given target (e.g., *candle*) compared to fluent instructions (see also Arnold, Fagnano, & Tanenhaus, 2003). These data suggest that disfluencies bias the addressee towards considering discourse-new objects as referents. Barr and Seyfeddinipur (2010) investigated this bias further in a mouse-tracking study using a female and a male voice and found that it was speaker-specific: The bias depended not just on the givenness from the listener's point of view, but on what was old and new for the current speaker. Arnold, Hudson Kam, and Tanenhaus (2007) showed that, in addition to the bias towards new objects, disfluencies can also lead to a bias to look at unfamiliar objects as compared to familiar objects. This effect was reduced when participants were told that the speaker suffered from object agnosia, suggesting that the familiarity effect reflected participants' inferences about the cause of disfluencies.

Building on previous findings that disfluencies are most likely to occur before a complex syntactic constituent (e.g., Hawkins, 1971), Bailey and Ferreira (2007) examined whether disfluencies are used as cues to disambiguate syntactic structures, They presented participants with sentences with disfluencies in different positions (e.g., "Put the uh uh apple on the towel in the box" vs. "Put the apple on the uh uh towel in the box"). Although participants correctly carried out the requested actions, they looked more at a towel which was placed in a box in the late than the early disfluency condition, reflecting the use of the disfluency for disambiguation of the syntactic structure.

In sum, the studies discussed previously are part of a general trend (also reflected in the work on common ground reviewed in the previous section) to go beyond scripted utterances and examine how people understand more natural utterances. So far the results show that listeners use aspects of the speech signal that were not traditionally viewed as relevant, such as disfluencies, as cues for core linguistic processes such as syntactic ambiguity resolution and the analysis of information structure.

3.4. Linguistic relativity

The relationship between language and cognition has been hotly debated for almost 100 years. Behaviorists (e.g., Watson, 1925) tended to believe that language and thought were essentially the same thing. Cross-linguistic study led linguists and anthropologists to propose that thought is determined by language-specific factors (see Gumperz & Levinson, 1996; Lucy, 1992; Sapir, 1921; Whorf, 1956). In contrast, cognitive scientists have typically assumed that language is distinct from perception and cognition (e.g. Chomsky, 1957, 1965). More recently researchers have started to investigate the issue experimentally. In the domain of color, for example, some evidence has accumulated that language is recruited involuntarily during simple perceptual tasks (e.g., Davidoff, Davies, & Roberson, 1999; Gilbert, Regier, Kay, & Ivry, 2006; Roberson, Davies, & Davidoff, 2000; Tan et al., 2008; Winawer et al., 2007). In the domain of motion, in contrast, no language-on-cognition effects could be found (cf. Gennari, Sloman, Malt, & Fitch, 2002; Papafragou, Massey, & Gleitman, 2002, 2006). The visual world paradigm is a promising experimental approach to investigate this issue.

Papafragou, Hulbert, and Trueswell (2008) investigated how native speakers of Greek and English allocate attention as they viewed motion events. Greek and English are languages that differ in the encoding of bounded motion: Greek speakers tend to use path verbs (e.g., ascend and cross), but English speakers predominately use manner verbs (e.g., slide, walk, and run). Papafragou et al. (2008) recorded the eye movements of participants viewing motion events during two tasks, a verbal description task and a memory (free-viewing) task. During the verbal description task significant cross-linguistic differences were found. Speakers gazed at the event components typically encoded in their language. However, during the memory (free-viewing) task no differences in eye gaze were observed. Papafragou et al. (2008) concluded that language-on-cognition effects arise only when language is recruited to achieve the task but not during event perception in general.

Papafragou and Trueswell (2010) compared eye movements of native speakers of Greek and English under conditions of cognitive load. When event encoding was difficult because participants were required to engage in a concurrent non-linguistic task (tapping), they looked longer at event components characteristically encoded in their language (i.e. English speakers preferred to inspect the path endpoint rather than the manner of motion region whereas Greek speakers showed the opposite pattern). This language-specific effect was absent when there was no concurrent task or when the concurrent task required the use of language (counting aloud). In a second experiment the concurrent task (counting aloud) was delayed (i.e. participants had 3 s to freely inspect the motion event before they were cued to start counting) and language-specific effects appeared just before engaging in the task (but not before). Overall, these data suggest that language-specific effects on attention emerge only in few situations and are related to how difficult participants perceive the task to be. Papafragou and Trueswell (2010) concluded that language can be recruited optionally for encoding events but that language-specific factors do not shape core biases in event perception and memory.

Huettig, Chen, Bowerman, and Majid (2010) explored similar issues by investigating how Mandarin numeral classifiers, a grammatical category in that language, influence listeners' eye gaze. If language-specific classifier categories influence processing, then on hearing a target noun participants should shift overt attention to objects that share the same classifier even when the classifier is not explicitly present in the speech stream. For example, on hearing the Mandarin word for scissors, they should look more at a picture of a chair than at a picture of an unrelated object because the nouns for scissors and chair share the classifier *ba3*. When Mandarin speakers heard a sentence that included a classifier (e.g., "ba3 scissors") they looked significantly more often at classifier-match objects (e.g., chair) than at distractor objects. However, when the classifier was not present in the spoken sentence (e.g., "scissors"), classifier-match objects (e.g., chair) were not more likely to be fixated than distractor objects. This demonstrates that classifier distinctions influence eye-gaze behavior, but only when classifiers are present in the speech stream. This study therefore suggests that language-specific effects on visual attention only arise when the language-specific distinction is *being produced or comprehended*, but not necessarily in cognition more generally (cf. Slobin, 1996, 2003).

4. Studies of language processing at the word level

A great deal of visual world research has focused on lexical processing. Main areas of interest are phonological processing and processing of fine phonetic detail, bilingual word recognition, effects of context on spoken word recognition, word-level semantic and conceptual processing, and the levels of representation at which visually derived representations are matched with language-derived representations during language–vision interactions.

4.1. Phonological/phonetic processing

Key issues in this area are the time-course and the dynamics of spoken word recognition, and the processing of phonemic cues and phonetic detail. For instance, Allopenna et al. (1998) investigated whether during spoken word recognition potential lexical candidates with mismatching onsets are activated (i.e. whether on hearing "beaker", words such as "speaker" are also activated). It has long been known that lexical candidates with word-initial phonological overlap (cohort competitors) compete strongly for recognition (e.g., on hearing the spoken sequence /kæp../ all words that start with these sounds, such as captain and captive, are activated in parallel, Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). However, continuous mapping models of spoken word recognition (e.g., TRACE, McClelland & Elman, 1986, and Shortlist, Norris, 1994) assume that lexical access is continuous and thus predict that rhyming words (e.g., "beaker"/"speaker") should also be at least weakly activated. The evidence for such rhyme competitor effects from reaction time studies is inconclusive (Connine, Blasko, & Titone, 1993; Marslen-Wilson & Zwitserlood, 1989; Shillcock, 1990). In their visual world study Allopenna et al. asked participants to "Pick up the beaker. Now put it …" in the context of a visual display of objects including, for instance, a beaker, a beetle (a phonological onset overlap competitor), a speaker (a phonological rhyme competitor), and a carriage (an unrelated distractor). They found that the likelihood of fixations to both the beaker and the beetle increased as the word "beaker" was

heard. As acoustic information from "beaker" started to mismatch phonologically with "beetle", the likelihood of looks to the beetle decreased as the likelihood of looks to the beaker continued to rise. Looks to "speaker" started to increase as the end of the word "beaker" unfolded. Thus, onset competitors of the target competed earlier (and more strongly) for overt attention than rhyme competitors (see also Magnuson, Tanenhaus, Aslin, & Dahan, 2003, for similar results using an artificial lexicon). These data confirm that acoustic information at the beginning of spoken words is more important than acoustic information later on in the word, but they also suggest that onset-mismatching phonological overlap nevertheless constrains lexical selection, as predicted by TRACE (McClelland & Elman, 1986) and Shortlist B (Norris & McQueen, 2008).

Magnuson, Dixon, Tanenhaus, and Aslin (2007) investigated the effects of word frequency (see also Dahan, Magnuson, & Tanenhaus, 2001), cohort density (i.e. the number of words overlapping at word onset), and neighborhood density (i.e. the number of words that mismatch with the target by only one phoneme at any word position) on target object fixations. They displayed target objects with three unrelated distractors; competitors were not displayed. High frequency targets were fixated more and earlier than low frequency targets. Targets with high cohort density were fixated less often and later than targets with a low cohort density. There was a crossover effect for neighborhood density: early on during the word, targets with a high neighborhood density were fixated more than targets with a low neighborhood density but after word offset this effect was reversed. These data suggest that neighborhood cohorts and recognition cohorts are not static but that competitor sets change dynamically over time.

Some visual world studies have investigated how listeners deal with the effects of connected-speech and casual speech processes. Conversational speech contains many phonological reductions (i.e., words are pronounced with fewer or different phonemes than in their canonical form). Mitterer and McQueen (2009) presented visual displays which contained the printed Dutch words "tas" (bag) or "tast" (touch) and two distractors. Listeners fixated the /t/-final words (e.g., tast) more often when the spoken Dutch sentence ("...tas") continued with the Dutch word "boven" (above) than when it continued with the word "naast" (next). Mitterer and McQueen (2009) argued that this behavior reflected Dutch speech production because word-final /t/ in Dutch is typically reduced more before /b/ than before /n/. They concluded that listeners use probabilistic knowledge about the effect of following context pre-lexically to resolve lexical ambiguities caused by continuous-speech processes. Brouwer (2010) used the printed-word version of the paradigm to examine whether spoken word recognition in casual conversational speech (containing many speech reductions) differs from carefully articulated laboratory speech (as used in almost all psycholinguistic – including visual world – experiments). Brouwer compared the recognition of canonical forms of mid-to-high frequency content words in displays of four words, one of which was the target. Brouwer constructed canonical form competitors (e.g., "companion" for "computer" which phonologically overlapped more at onset with the canonical form than with the reduced form of the spoken word) and reduced form competitors (e.g., "pupil" for "puter"), which phonologically overlapped more at onset with the reduced form than with the canonical form of the spoken word ("computer"). Listeners directed significantly more overt attention to the canonical form competitor than to the reduced form competitor in both a laboratory speech condition and a casual speech condition when the spoken sentences contained no speech reductions. However, when the speech contained reductions there was no difference between listeners' fixations to canonical form and reduced form competitors. Brouwer concluded that during casual speech, which includes a great deal of reduced word forms, listeners are more tolerant of acoustic mismatches between input and canonical form than in more formal speech. These data therefore show that speech-intrinsic variation (e.g., the overall reliability and quality of the phonetic input) can influence phonological competition.

Several other studies have investigated processing of phonetic detail. Reinisch, Jesse, and McQueen (2010) found that visual world participants can use lexical stress information to direct eye gaze. They found that when participants heard words with initial stress (e.g., "octopus") fixations on printed target words with stress on first syllable (e.g., octopus) were more frequent than fixations on differently stressed competitors (e.g., October, with stress on second syllable) before segmental information could disambiguate the words. The authors concluded that listeners recognize words by immediately using all relevant information in the acoustic signal. McMurray and colleagues used the visual world paradigm to show that spoken word recognition exhibits graded sensitivity to within-category voice onset time (a strong cue to distinguish voiced sounds such as /b/ from voiceless sounds such as /p/; McMurray, Tanenhaus, & Aslin, 2002; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2009). Salverda, Dahan, and McQueen (2003) varied the duration of an ambiguous acoustic sequence (see also Shatzman & McQueen, 2006). In one of their displays, they presented pictures of a piece of ham, a hamster, and two unrelated distractors. They observed that participants looked more at the picture of the ham when the first syllable of the target word (i.e. "hamster") stemmed from a recording of the monosyllabic word "ham" than when it stemmed from a different normal recording of "hamster". Salverda et al. found that this effect was due to the fact that the acoustic sequence "ham" is longer in the monosyllabic word "ham" than in "hamster". They concluded that listeners can use fine phonetic detail such as segmental lengthening during spoken word recognition.

These results demonstrate the listeners' sensitivity to phonetic detail in the utterances they hear. It is unclear however to which extent the effects were driven by prior activation of lexical candidates from viewing the objects (e.g., from viewing the ham and the hamster in the study by Salverda et al.). Perhaps subtle cues in the speech signal can be used in visual world studies because of such pre-activation but in the absence of strong visual support such cues are much less likely to be used (but see Tanenhaus, Magnuson, Dahan, & Chambers, 2000, for a different view).

The situation is probably different for *strong* cues, such as those signaling the beginning of a word, which are likely to be important even in the absence of priming from pictures. The study by Allopenna et al. (1998) illustrates this point. In contrast to robust onset overlap effects, rhyme competitor effects (beaker/speaker) are small and typically only marginally statistically significant (see also Allopenna et al., 1998; Huettig & McQueen, 2009; McQueen & Viebahn, 2007). It is thus unclear whether rhyme competitors play a significant role during speech processing in the more common situations in which there is no pre-activation of candidate words. As mentioned previously, prior to the demonstration of rhyme competitor effects in spoken word recognition by Allopenna et al. (1998), such effects had been difficult to demonstrate using methods such as cross-modal or auditory–auditory priming. Connine et al. (1993) and Andruski, Blumstein, and Burton (1994) found weak priming of onset-mismatching items only when prime and target differed by no more than one or two phonetic features. Allopenna et al. (1998), in contrast, found consideration of the rhyme competitor even though it differed by more than two features from the target word. It is possible that the visual world method is more sensitive to detect rhyme competitor effects because of the pre-activation of the rhyme competitors via preview of the visual objects.

There are situations in which weak cues can become particularly important, for example when, in noisy listening conditions, strong onset cues become less reliable. Huettig and McQueen (2009) recently demonstrated how environmental noise can increase the influence of rhyme overlap during spoken word recognition. They

replicated the results obtained by Allopenna et al. with participants fixating onset-overlap competitors more than rhyme-overlap competitors, but the strength of this tendency varied with speech quality. Relative to a baseline with noise-free sentences, participants looked less at onset-overlap and more at rhyme-overlap pictures when some phonemes in the sentences (but not in the critical words) were replaced by AM-radio noise. The position of the noise in the surrounding words (word-initial or word-medial) had no effect. Thus noise elsewhere in the sentences apparently made evidence about the critical word less reliable: Listeners became less confident that they had heard the onset-overlap name and that they had not heard the rhyme-overlap name. The same acoustic information therefore has different effects on phonological competition during spoken-word recognition as the probability of distortion in the environment changes.

In sum, visual world research on phonological and phonetic processing has shown that subtle phonetic cues *can* modulate lexical activation. However, it is possible that the effects reported in the studies were at least partly driven by the pre-activation of the lexical candidates from viewing the stimulus display. This is particularly likely in the printed word version of the paradigm because printed words readily activate the corresponding phonological forms (Van Orden, Johnston, & Hale, 1988; see also Frost, 1998). With the competitors 'in mind', it might become easier to attend to the specific cues in the speech signal which disambiguate between the different items in the display very early. It is conceivable that in other listening conditions and conversational situations subtle cues play a negligible role.

4.2. Bilingual word recognition

An important issue in bilingualism research is the question whether lexical access in bilinguals is language-specific (i.e. restricted to the intended language) or alternatively both languages are active and may influence performance. A seminal study of this issue was conducted by Spivey and Marian (1999). They presented Russian–English bilinguals with Russian sentences such as "Poloji marku nije krestika" ("Put the stamp below the cross"). There were four objects in the visual display: the stamp, a marker, and two unrelated distractors. Critically, the English phonological form "marker" is phonologically similar to the Russian word "marku" (stamp). On hearing the Russian word "marku" participants looked more often at the picture of a marker than at distractor objects though the Russian word for marker shares no phonological similarity with the spoken word. Spivey and Marian (1999) concluded that bilingual listeners cannot deactivate their other language when in a monolingual situation (see also Marian & Spivey, 2003a, 2003b).

However, later visual world studies suggest that this conclusion might be too strong. Weber and Cutler (2004) could not find any evidence that Dutch listeners activate English words when listening to Dutch sentences (i.e. when hearing the Dutch word "deksel", lid, they did not fixate the picture of a desk more than unrelated distractor pictures). It should be noted that the samples used in the two studies were quite different. The Russian–English bilinguals in the Spivey and Marian (1999) study had immigrated to the US as teenagers, were studying at a top-tier US university, and were completely immersed in English in their daily lives. By contrast, the Dutch–English bilinguals in Weber and Cutler's study (2004) lived in the Netherlands and mostly used Dutch in everyday life.

Weber and Cutler (2004) also presented Dutch listeners and native British–English listeners with spoken English sentences. Participants were required to click on items in a display containing, for instance, a panda and a pencil. Note that Dutch speakers have difficulty distinguishing English words containing /æ/ from words containing /ɛ/. Weber and Cutler found that Dutch listeners hearing "panda" were more likely to look at the pencil than at unrelated distractors. English listeners did not do this. Interestingly, when asked to "click on the pencil", Dutch listeners did not show increased fixations to the panda (i.e. their performance was similar that of the native English listeners). Weber and Cutler suggested that the asymmetric interference effect in their study arose because at the phonetic processing level one of the L2 categories is dominant (i.e. /ɛ/). Dominance appears to be determined by acoustic–phonetic proximity to the nearest L1 category (Cutler, Weber, & Otake, 2006). Therefore at the lexical processing level, representations containing this dominant category are more likely than representations containing the non-dominant category (i.e. /æ/, which does not exist in Dutch) to be contacted by the phonetic input.

Data obtained by Ju and Luce (2004) shed further light on the influence of phonetic/phonological information for bilingual listeners. They failed to replicate the findings by Spivey and Marian (1999) using unaltered Spanish targets. Ju and Luce (2004) observed interference only when the Spanish target words contained English-appropriate voice onset times. Ju and Luce (2004) concluded that fine-grained acoustic–phonetic information and a precise match between acoustic input and stored representations are critical for parallel activation of two languages (see also Pallier, Colome, & Sebastian-Galles, 2001; Li, 1996; Grosjean, 1988).

Canseco-Gonzalez et al. (2010) had English–Spanish bilinguals follow instructions such as "Click on the beans". Cross-linguistic competitors had a similar onset phoneme in Spanish (e.g., 'bigote' and mustache). Canseco-Gonzalez et al. (2010) found a weak cross-language effect, which was modulated by the age of acquisition of Spanish (with the early bilinguals showing little evidence of cross-linguistic activation). Their findings suggest that the degree of cross-linguistic activation is influenced by the age of acquisition of each of the languages.

In sum, it appears that visual context, fine-grained acoustic information, age of acquisition, language proficiency, and mode of processing all determine to what extent bilinguals activate representations of their other language when in a monolingual situation (see the special issue of *Acta Psychologica*, Volume 128(3), 2008, for detailed discussion of current bilingual research).

4.3. Influence of semantic and syntactic context on spoken word recognition

The issue of modularity (discussed in Section 3 on sentence processing) has also influenced much of the psycholinguistic work on lexical (i.e. word) processing. Many classical studies using a variety of paradigms suggest that multiple meanings of ambiguous words are activated even when semantic or syntactic constraints should induce a strong bias towards one of the meanings (e.g., Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979; Whitney, McKay, Kellas, & Emerson, 1985). These studies suggest that there is at least some initial bottom-up priority when listeners perceive the acoustic signal (cf. Marslen-Wilson, 1987, 1990). Other studies have found that context can have very early effects on the activation of word meanings (Moss & Marslen-Wilson, 1993; Tabossi, 1988; Tabossi, Colombo, & Job, 1987; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003).

In a visual world study, Dahan and Tanenhaus (2004) presented their Dutch participants with visual displays of four objects: a target object (e.g., a goat, Dutch "bok"), a phonological competitor (e.g., a bone, Dutch "bot"), a semantic competitor (e.g. spider, Dutch "spin") and an unrelated distractor. There were two conditions. In the constraining-verb condition a contextually constraining verb (e.g., climb) appeared before the critical noun (e.g., "Nog nooit klom een bok zo hoog" — Never before climbed a goat so high). The semantic competitor (e.g. spider, Dutch "spin") was selected to be a plausible subject of the verb but phonologically different from the target (goat, "bok"). In the neutral condition the constraining verb followed the

critical noun (e.g., "Nog nooit is een bok zo hoog geklommen" — Never before has a goat climbed so high). In both conditions there was a small semantic competitor effect (i.e. more looks to the spider than the unrelated distractor). In the neutral condition participants looked more at both the goat (i.e. bok) and the bone (i.e. bot) than at the phonologically unrelated distractors as the target word was heard. This result reflects the standard competition effect of phonological forms overlapping at word onset (e.g., Allopenna et al., 1998). By contrast, in the constraining verb condition participants did not look more at bot (bone) than at the unrelated distractors on hearing "bok", suggesting an immediate effect of context on word recognition.

The absence of looks to the phonologically related competitor in the constraining condition does not necessarily mean that the word was not activated. Participants in the visual world paradigm try to make sense of the speech and the visual input to interpret the situation at hand. In the Dahan and Tanenhaus (2004) study, participants encountered the biasing verb "climb" and anticipated that bok (the goat) was the intended referent (cf. Altmann & Kamide, 1999) as evident from increased fixations to bok (goat) even before acoustic information from the critical noun became available. In order to separate verb-based anticipatory effects from effects reflecting the immediate integration of phonetic/phonological information from the target word with the semantic context, Dahan and Tanenhaus (2004) separately analyzed the 10 items in the constraining verb condition which did not show an anticipation effect towards the target (i.e. these items showed an anticipation effect towards the semantic competitor). Analysis of these 10 items showed a similar pattern as the overall results: there was no shift towards the phonological competitors (i.e. no more looks to "bot", bone, than unrelated distractors when the verb biased "bok", goat). Dahan and Tanenhaus argued that their results "favor models in which mapping from the input onto meaning is continuous over models in which contextual effects follow access of an initial form-based competitor set" (p. 498; see also Magnuson, Tanenhaus, & Aslin, 2008, for a similar interpretation of a null effect).

The results are, however, open to a different interpretation. The phonological representations of bot (bone) may be just as strongly activated by the acoustic signal as in the neutral condition because eye gaze is a measure of overt attention and not of underlying linguistic representations. It cannot be ruled out that overt attention reflects the *outcome* of a process which integrates context with initially exhaustive (i.e. form-based) lexical access (i.e. that at an early stage during the processing of the target all related words, regardless of their contextual fit, were activated).

4.4. Mapping language-derived and vision-derived representations

A different line of research investigates how representations accessed on hearing spoken words are integrated with the conceptual and perceptual information accessed from viewing visual scenes or displays of objects. The main theoretical issues of this line of research are to understand (i) whether language-mediated eye movements are driven by similarity between targets and competitors or by all-or-none categorical knowledge, (ii) whether they are mediated by perceptual properties or stored knowledge about the depicted objects, (iii) whether some types of knowledge (e.g., functional information) are prioritized over others, and (iv) how these mapping processes are affected by cognitive control and situational demands.

Cooper (1974), in his seminal visual world study, observed that participants hearing "Africa" were more likely to fixate pictures showing a lion, a zebra, or a snake than semantically unrelated objects. Thus, the eye movements revealed the on-line activation of word semantics from the speech input. Cooper (1974) did not investigate systematically the nature of the semantic effects he observed. For instance, the words "Africa" and "lion" are associatively related and so it is unclear whether Cooper's semantic effects were driven by semantic similarity or by mere association.

Following up on the work of Cooper (1974), Huettig and Altmann (2005) found that participants directed overt attention towards a depicted object (such as a trumpet) when a semantically related but not associatively related target word (e.g., "piano") was heard (see also Yee & Sedivy, 2006; Dunabeitia, Aviles, Afonso, Scheepers, & Carreiras, 2009). Importantly, the probability of fixating the semantic competitor correlated significantly with the semantic similarity between the spoken word (e.g., "piano") and competitor object (e.g., trumpet) as derived from semantic feature norms (Cree & McRae, 2003). These data suggest that the increased attention directed to semantically related items (relative to distractor objects) was a function of the degree of semantic overlap.

Huettig, Quinlan, McDonald and Altmann (2006) (see also Yee, Overton & Thompson-Schill, 2009) provided further evidence for this conclusion. They found that several corpus-based measures of word semantics (Latent Semantic Analysis, Landauer & Dumais, 1997; Contextual Similarity, McDonald, 2000) each correlated well with fixation behavior. These data provide strong evidence that language-mediated eye movements are driven by semantic similarity between the spoken word and the visual object rather than by all-or-none categorical knowledge.

Given the observed effects of semantic/conceptual overlap, the question arises whether *perceptual* overlap between a target and a distractor affects language-mediated eye movements. Several studies have found that eye gaze is directed to objects that are visually related (i.e., by shape) to the targets but are semantically unrelated (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004, 2007). For example, when hearing "snake" participants shifted overt attention to a picture of a cable (cable and snake have a similar global shape), even though they had viewed the four-object display for approximately 5 s before hearing the target word, which was sufficient time to recognize the objects for what they were (Huettig & Altmann, 2007). Dahan and Tanenhaus (2005) interpreted these results as demonstrating that the probability of fixating visual object reflects a match between stored visual-form knowledge, which is accessed from the spoken word, and visual features accessed from the visual display (a "coarse structural representation associated with each object's location" (Dahan & Tanenhaus, 2005, p. 457). In a related study, Huettig and Altmann (2004) found that eye movements in the visual world paradigm could be mediated by color relations: Participants shifted overt attention to a picture of a strawberry when they heard "lips"; strawberries and lips have a similar typical color.

An interesting question raised by these findings is whether the participants' eye movements were affected by the visual input (i.e. the perceptual properties) or by the stored knowledge about the depicted objects. The shape of an object, the long and thin form of a snake for example, can be *perceived* (perceptual information) but is also *known* (conceptual information). Therefore, it is unclear whether the shape-driven shifts in overt attention were caused by the stored (conceptual) knowledge of the shape of the displayed objects or by the perceived shape of the objects in the visual display (see also Yee, Huffstetler, & Thompson-Schill, 2009). Similarly, it is unclear whether the color effect was contingent upon the *stored* knowledge about the typical color of the displayed object or whether language-mediated attention can also be contingent upon the *perceived* surface color of the visual objects.

Color is an object property that allows for the investigation of this question since conceptual attributes (the *stored* color knowledge about an object) and perceptual attributes (the *perceived* but non-diagnostic color of an object, i.e. its surface color) can be dissociated. Huettig and Altmann (2011) presented participants with spoken target words whose concepts are associated with a diagnostic color (e.g., "spinach", spinach is typically green) while their eye movements were monitored to (i) objects associated with a diagnostic color but presented in black and white (e.g., a black and white line drawing of a frog), (ii) objects associated with a diagnostic color but presented in

an appropriate but atypical color (e.g., a color photograph of a yellow frog) and (iii) objects not associated with a diagnostic color but presented in the diagnostic color of the target concept (e.g., a green blouse). Huettig and Altmann found no effect of stored object color knowledge when black and white line drawings or black and white photos were presented. A weak effect of stored object color knowledge was observed when color photographs were used depicting the target object (e.g., a frog) in an atypical but appropriate color (e.g., a yellow frog). This suggests that participants retrieved prototypical color information (i.e., green) on hearing "spinach", which matched with the stored prototypical color information they retrieved from seeing the yellow frog (i.e., that frogs are typically green). However, the effect was marginal and occurred rather late (more than one second after information from the acoustic target word started to become available). These results suggest that stored object color can have an influence on language-mediated eye movements but this influence appears to be small.

The finding that there was no such effect with black and white stimuli suggests that the absence of color in the stimuli induced an attentional bias that resulted in the null effect of color properties. It has been argued in the visual search literature that effects of a particular visual feature may be more or less likely to draw attention depending on the attentional control setting adopted by the participant (Folk, Remington, & Johnston, 1992, 1993; see also Pratt & Hommel, 2003). Similarly, according to dimension-based theories of selective attention (e.g., Allport, 1971; Müller, Heller, & Ziegler, 1995), top-down information of target defining features is assumed to influence feature-processing stages (Found & Müller, 1996; Treisman, 1988; see also Kumada, 2001). It is thus possible that the black and white stimuli reduced attention to the color dimension because it was of little relevance in the particular context.

Huettig and Altmann (2011) also investigated the effect of the perceived surface color in absence of stored color knowledge. On hearing target words that were associated with a prototypical color, such as "pea", participants looked towards a picture displayed in that color even though the referent of the picture (e.g., a green blouse) was not itself associated with that color. On accessing the prototypical color information of the target referent, participants shifted overt attention immediately to anything with the same surface color. These data are a clear demonstration of a pure surface color effect. Overall, these experiments reveal that color-mediated shifts in overt attention are primarily due to the perceived surface attributes of the visual objects rather than stored knowledge about the typical color of the objects.

It is however possible that the color effects are nevertheless mediated by stored color labels (e.g., the label "green" accessed on hearing the spoken words). When participants hear the spoken word "frog", they may automatically access the word "green", which may then cause them to shift their eye gaze to anything that is green in the display. This notion is supported by responses in free word association tasks. When participants are asked to write down the first word they think of when reading the word *frog*, they typically give the response *green* (e.g., Nelson, McEvoy, & Schreiber, 1998). Johnson, Huettig and McQueen (2008) investigated this issue with young children who have yet to learn any color terms. They pre-tested two-year-olds on their understanding of color terms (e.g., "point to the red thing"). They found that toddlers who did not understand color labels showed similar color-mediated eye movements to adults. For example, on being asked to find a strawberry, they looked more at a red plane than a yellow plane despite being unable to name the colors. These results suggest that the color effect in the visual world paradigm, at least in young children, is not mediated by color labels.

Finally, it appears that some types of knowledge are prioritized by visual world participants. Categorical and functional knowledge for instance are particularly salient aspects of lexical knowledge (see Moss, McCormick, & Tyler, 1997). Huettig and Altmann (2011) found that gazes to a surface color competitor occurred 400 ms later in a condition in which a semantic competitor was co-present in the display than in a condition in which the surface color competitor was the only related object in the display. These data also illustrate an important aspect of the paradigm, namely that the time-course of gaze to targets and competitors in visual world studies must be interpreted with caution. The fact that the presence of another (e.g., semantic) competitor alters the timing by which a surface color competitor attracts overt attention shows that shifts in eye gaze are, at least in some cases, not a direct reflection of activation of corresponding representations. The finding that surface color competitors attracted attention 400 ms later when semantic competitors were co-present than when they were not co-present, suggests that (even for natural objects that are relatively high in color diagnosticity) categorical/functional knowledge is a more salient aspect of knowledge than color knowledge and is prioritized by the attentional system.

The studies discussed previously showed that listeners in the visual world paradigm use visual as well as semantic and conceptual information about the objects when they integrate the visual and the auditory input. There is, however, evidence from other paradigms showing that viewers often access the names of objects, even when they do not intend to name them (e.g., Morsella & Miozzo, 2002; see also Meyer & Damian, 2007; Meyer, Belke, Telling, & Humphreys, 2007; Navarette & Costa, 2005). Thus, it is conceivable that the objects shown in visual world studies also activate their names, and that this lexical information may affect the participants' mapping between the visual and the spoken input. Results obtained by Huettig and McQueen (2007) support this suggestion. In their Experiment 1, participants saw visual objects for about three seconds before target onset. Participants shifted their eye gaze to objects (e.g., a beaver) whose phonological form had the same word-initial phonemes as the spoken target word (e.g., "beaker") *before* they looked at objects which were semantically (e.g., a fork) or visually (e.g., a bobbin) similar to the target word. The time precedence of fixations to the phonological competitors suggests that participants accessed the names of the pictures before they heard the critical spoken words. Moreover, research on spoken word recognition strongly suggests that spoken word recognition is a cascaded process (see McQueen, Dahan, & Cutler, 2003, for review). On hearing the spoken target word different candidate words compete for recognition in parallel at the phonological level of representation if they are consistent with the acoustic information in the speech signal. Processing at the phonological level does not have to be completed before information cascades to other (e.g., semantic or visual-form) levels of representation. Some phonological processing however does precede processing at other levels of representation because of the primacy of acoustic–phonetic information in the speech signal. Huettig and McQueen (2007) argued that this is why fixations to phonological competitors preceded shifts in eye gaze to semantic and visual-form competitors: phonological information was accessed first from the spoken words and could be used first in the mapping process between the spoken word and the visual objects. Huettig and McQueen's Experiment 2 provides further support for this account. Participants were presented with the visual display for only 200 ms before the onset of the spoken word. Now the participants did not fixate phonological competitors more than unrelated distractors. Huettig and McQueen (2007) argued that this was because 200 ms was not enough time to retrieve picture names before the phonological competitor ceased to be a viable lexical hypothesis. In other words, when there was no time to retrieve picture names there were no fixations to the phonological competitors. This strongly supports the notion that the looks to the phonological competitors in Experiment 1 were indeed due to mapping of phonological representations.

In sum, the studies reviewed in this section suggest that participants can retrieve phonological, semantic, and visual-form information from objects in the visual displays and from the spoken words and any of these representations may be important to the

process of integrating linguistic and non-linguistic information. Attentional shifts thus appear to be co-determined by the type of information in the display (i.e., pictures or printed words), the timing of cascaded processing in the word and picture recognition systems, the temporal unfolding of information in the speech signal, the number (and type) of competitors which are co-present in the display, and the attentional setting of the participant.

5. Production studies using the visual world paradigm

The visual world paradigm has primarily been used to study spoken language comprehension. However, there are a number of production studies that have used a related methodology. Here we review three lines of production research: studies of message generation, of utterance formulation, and of self-monitoring.

5.1. Message generation

An important and largely unresolved issue in current speech production research is how speakers generate preverbal messages. These are the conceptual structures that determine the utterance content and form the input to the linguistic encoding processes (e.g., Bock, Irwin, & Davidson, 2004). Studying message generation experimentally is taxing because in everyday language use, message generation – which is, essentially, deciding what to say when – depends on many social, motivational, cognitive and linguistic factors, which are difficult to separate and manipulate. However, there are important aspects of message generation that can be studied by asking speakers to describe what they see and to record both their speech and their eye movements. This allows one to trace when speakers acquire the visual raw materials for the generation of a message and to relate this information to the timing, content and form of their utterances.

To our knowledge, Griffin and Bock (2000) were the first to use eye tracking to study message generation. They asked participants to describe cartoons of events (e.g., of a man chasing a dog) and compared their eye movements to those in various control tasks, including a patient-detection task, where the participants had to identify the event character who was undergoing (rather than carrying out) the action. Griffin and Bock found that initially the gaze patterns in the two tasks were very similar, but that after about 300 ms the likelihood of fixations on the agent and patient began to differ between the tasks: In the description task speakers first looked at the agent, whom they usually mentioned first, and then, shortly before mentioning it, turned to the patient. By contrast, in the patient-detection task, they turned to the patient and rarely looked at the agent. Griffin and Bock concluded that the speakers' gaze pattern was indicative of the existence of two distinct phases in their processing of the pictures and the spoken utterances: an apprehension phase of about 300 ms, during which they comprehend the event, and a following formulation phase during which they generate the linguistic form of the utterance. During the latter phase, speakers look at each of the objects they name in the order of mention.

The distinction between an apprehension and a formulation phase is supported by the results of a study by Bock, Irwin, Davidson, and Levelt (2003), who investigated message generation and linguistic formulation in time-telling utterances. They presented Dutch and English speakers with analogue and digital clock displays and asked them to produce absolute expressions (e.g., "two fifty") or relative expressions (e.g., "ten to three"), thereby creating situations where the required linguistic form was more or less compatible with the display (analogue displays are more compatible with relative expressions; digital ones with absolute expressions) and with the speakers' time telling preferences (Dutch speakers prefer relative expressions, whereas speakers of American English prefer absolute expressions). There were effects of compatibility and display type, but, most importantly, from about 300 ms onwards, the speakers' eye gaze depended on the type of expression. When speakers produced relative expressions ("ten to three") they looked first at the minute hand of the analogue display or the right part of the digital display, and then at the hour hand or the left part of the digital display. The opposite pattern was seen for absolute expressions ("two fifty"). Bock and colleagues conclude that "an effective interface between what has been seen and what is to be said, can be constructed within 300 ms. This interface underpins a preverbal plan or message that appears to guide a comparatively slow, strongly incremental formulation of phrases" (p. 653).

Gleitman et al. (2007) also asked participants to describe cartoons of actions. On most trials of the study, a brief visual cue appeared prior to picture onset in the position of one of the scene characters. Although the participants could not consciously perceive the cues, their first fixations were more likely to be directed at cued than at uncued characters. In addition, the cues affected the way the speakers described the events: The cues increased the likelihood of the cued character being mentioned first in the utterance, and, related to this, altered the speakers' choice of verb (e.g., "to chase" vs. "to flee") or their choice of active vs. passive structure. Gleitman and colleagues proposed that the cue captured the speakers' visual attention, and that the direction of visual attention to the cued character facilitated the retrieval of the character's name, which in turn increased its likelihood of being mentioned early. These results are entirely compatible with the proposal made by Bock and colleagues that speakers rapidly apprehend the gist of a scene, and that this early representation largely determines in which order different regions of the scene will later be inspected and mentioned. However, the influence of the visual cues on the direction of the early fixations and on the utterance form shows that it is difficult to draw a clear temporal distinction between a wholistic scene apprehension phase and a subsequent incremental linguistic encoding phase.

The interplay between message generation, specifically the uptake of new information, and grammatical encoding was studied by Brown-Schmidt and Tanenhaus (2006). They used displays showing several objects and a referential communication task, where participants produced noun phrases such as "the small triangle" or "the square with the small triangles". Using a size adjective was only necessary when the display included a contrast object that only differed in size from the target. The authors found that speakers were far more likely to use an adjective when they had fixated upon the contrast object than when they had not done so. More importantly, the timing of the fixations on the contrast object was related to the fluency of the utterance. Fluent utterances including prenominal adjectives were preceded by earlier fixations to the contrast object than utterances including repairs, such as "the square ... small one". Thus, the timing of the speakers' visual information uptake had direct consequences for the fluency and form of their utterances. These results highlight the incremental nature of utterance generation, i.e. the fact that message fragments corresponding roughly to individual words (rather than, for instance, entire phrases or sentences) appear to be passed on to the formulator (see also Brown-Schmidt & Konopka, 2008).

5.2. Utterance formulation

In a second line of production research, the perception of the display and the generation of the message are deliberately simplified as much as possible, usually by asking participants to name sets of objects in a fixed order, and eye tracking is used to study the time course of lexical access in multi-word utterances, for instance how far ahead speakers plan their utterances and whether they can retrieve several object names in parallel. Lexical access is often investigated in paradigms measuring speech onset latencies, but latencies only provide evidence about the time speakers need to plan the first word of an utterance. By contrast, eye movements can be recorded both before and after speech onset and, as will be shown later, can provide important additional information about the way

speakers coordinate visual information uptake, linguistic planning, and articulation with each other.

This line of research has uncovered tight links between the speakers' eye gaze and their overt speech output. First, as in the studies of message generation, speakers typically look at each object they refer to shortly before mentioning it (Griffin, 2001; Meyer et al., 1998). This is true even when they name the same objects on many successive trials (Wheeldon, Meyer, & van der Meulen, 2007). Second, the order of the inspection of the objects closely corresponds to the order of mention, and when speakers talk about the same object twice within a short utterance (as in "The box next to the star is yellow"), they tend to inspect the object twice, shortly before the onset of the first word referring to the object ("box") and before the onset of the second word ("yellow"; Meyer, van der Meulen, & Brooks, 2004). Thus speakers strongly prefer to look at the objects they refer to even when they could easily retrieve the relevant information from working memory.

In addition, the timing of the speakers' eye movements has been shown to be tightly coordinated with their speech planning. Studies varying the ease of identifying the objects, of selecting appropriate lexical items and of generating the corresponding sound forms have provided strong evidence that speakers typically fixate upon each object they name until they have recognized it and retrieved the phonological form of the referring expression (Belke & Meyer, 2007; Griffin, 2001; Levelt & Meyer, 2000; Meyer, Roelofs, & Levelt, 2003; Meyer & Van der Meulen, 2000). For instance, speakers look longer at objects with long names than at objects with short names (Meyer et al., 2003; but see Griffin, 2003), which indicates that the shift of gaze from one object to the next only occurs after the phonological form of the first object name has been retrieved.

Different proposals have been made concerning the origins of the tight temporal link between eye movements and speech planning (for further discussion see Griffin, 2004). In considering this issue it is important to keep in mind that, by and large, a person's point of gaze corresponds to the focus of their visual attention, and that directing one's visual attention to an object probably facilitates not only the identification of the object but also the retrieval of associated information, including the name of the object (e.g., Humphreys & Forde, 2001; Humphreys, Riddoch, & Price, 1997; Roelofs, 1992; see also Griffin, 2004; Wühr & Frings, 2008; Wühr & Waszak, 2003; Meyer, Belke et al., 2007; Meyer, Ouellet et al., 2008, for further discussion). A likely reason why speakers attend to the objects in the order of mention and as long as they do is that this facilitates object recognition, lexical retrieval and the internal self-monitoring processes that precede overt utterances. This proposal fits in well with recent evidence demonstrating that at least some components of lexical access and self-monitoring require processing capacity and would therefore benefit from attentional enhancement (e.g., Cook & Meyer, 2008; Ferreira & Pashler, 2002; Huettig & Hartsuiker, 2008, 2010; Roelofs, 2007, 2008).

The studies discussed so far suggest that speakers generate simple descriptive utterances in a highly incremental fashion, as they only turn to a new object when they are almost ready to initiate the name of the current object. However, the eye movement data may be somewhat misleading because the object a person is fixating upon need not be the only object they are attending to. Instead viewers can use a broader attentional focus and attend to several objects in parallel (e.g., Cave & Bichot, 1999). Several studies have investigated whether this indeed happens during speech production (Meyer, Ouellet & Häcker, 2008; Morgan & Meyer, 2005; Morgan, van Elswijk, & Meyer, 2008). These studies have yielded clear results: Speakers can attend to two objects – a foveated object they are about to name and an extrafoveal object they will name next – in parallel. Moreover, the processing of the extrafoveal object can be sufficient for its name to be activated in parallel with that of the foveated object. However, parallel object processing only occurs when both objects are relatively easy to recognize and name. For instance, Malpass and Meyer (2010) found that the first of two objects was fixated for a longer time and was named more slowly when the second object was easy than when it was difficult to name. This suggests that the two objects were processed in parallel, and that the easy-to-name second object rapidly activated its name, which interfered with the retrieval of the name of the first object. This interference effect disappeared when the visual processing of the first object was made more demanding by presenting it upside-down. Thus, parallel processing of two objects and parallel retrieval of their names is possible, under favorable conditions. Further research is required to determine the conditions under which it is likely to occur.

5.3. Self-monitoring of spoken words

Theories of speech monitoring (Hartsuiker & Kolk, 2001; Levelt, 1989; Postma, 2000) assume that speakers, in addition to hearing their own overt speech, can 'inspect' an internal representation of their planned speech before articulation (see Dell & Repka, 1992; Lackner & Tuller, 1979; Motley, Camden, & Baars, 1982; Oppenheim & Dell, 2008, for experimental evidence supporting this assumption). There are two theoretical accounts of how speakers monitor this internal representation. One influential account proposes that internal monitoring engages speech perception (Levelt, 1989), the other that it engages language production internal devices (e.g., Laver, 1980; see Postma, 2000, for review). Perception-based theories predict that listening to one's own inner speech has similar behavioral consequences as listening to someone else's speech. To test this view, Huettig and Hartsuiker (2010) registered eye-movements while speakers named objects accompanied by phonologically related or unrelated written distractor words. In the critical condition, one of the distractor words was phonologically related to the name of the target object and two words were unrelated. Huettig and Hartsuiker (2010) found that participants fixated the phonological competitors more than the unrelated distractors. Importantly, the time course of fixating the phonological competitors during self-perception was very similar to the time course observed when the participants listened to another speaker (Huettig & McQueen, 2007, Experiment 4) which suggests that the eye movements were driven by overt (and not inner) speech. This study suggests that external but not internal self-monitoring is based on speech perception. Huettig and Hartsuiker (2010) concluded that there is a need for more elaborated theories of the alternative viewpoint, namely production-internal monitoring. In addition, their finding that perceiving one's own speech during articulation can drive eye movements in a very similar way to listening to someone else's speech further highlights the tight coupling between language processing (self-produced or other-produced) and overt gaze.

In sum, the production studies, just like the comprehension studies, have revealed that the viewers' visual inspection of the displays is tightly coordinated with their linguistic processing. This is because in both cases, the eye movements reflect the direction of visual attention. Speakers and listeners use visual attention in order to cope efficiently with their respective tasks of producing utterances and of mapping the utterances they hear onto the visual arrays. That is, people carrying out linguistic tasks look at relevant objects not only to identify these objects, but also because looking – or, rather, attending – facilitates the retrieval of information about these objects (see also Griffin & Oppenheimer, 2006). It follows that eye tracking can be used to determine when and for how long speakers (and listeners) focus their attention on different parts of a display, though one should keep in mind that regions that are not fixated may nevertheless be processed and that therefore the onset of the gaze to a region cannot be equated with the onset of processing. It is worth keeping in mind that eye movements do not reveal why a person is attending to a region – whether it is, for instance, to see it clearly, or check whether an utterance they have already produced is correct; and they don't reveal why a gaze to an object is long or short – a gaze

might be long because the object is difficult to identify or because it is difficult for the speaker to retrieve a suitable name. To reiterate, eye gaze reflects visual attention, and any variable that affects when and for how long a viewer decides to attend to an object is likely to affect when and for how long they look at it as well.

6. Summary and conclusions

The comprehension version of the visual world paradigm is characterized by three defining features: (1) on each trial the participants hear a stretch of speech, (2) they also see a relevant visual display, and (3) their eye movements are recorded for later analyses. In the production version of the paradigm, participants are shown visual displays and are instructed in more or less specific ways to talk about them. As comprehension researchers can combine any visual display with any speech input, and similarly production researchers can elicit, through questions or instructions, an unlimited range of utterances, the paradigm is extremely versatile. Somewhat paradoxically perhaps, most of the visual world research has concerned language processing, but the dependent measures, saccades and fixations, concern the visual exploration of the displays. The reason why this approach has been successful is, of course, that the language users' eye movements are systematically related to their linguistic processing. Yet, the link between language processing and eye movements, though systematic and powerful, is indirect.

In comprehension studies the spoken words do not pull the eyes to certain objects on the screen; instead, the speech-eye link arises because the verbal information affects the listeners' allocation of attention, which in turn governs the direction of their gaze. As explained previously, in some visual world studies, the verbal input constituted instructions to the participants to look at or move objects in the display, and it is not surprising that they indeed fixated upon those objects. In other studies, where no specific instructions about the purpose of the sentences and pictures were given, the participants probably interpreted the spoken utterances as a commentary on the displays, or viewed the displays as illustrations of the spoken utterances, and aimed to create mental representations that linked the two types of information in a meaningful way. As we argued previously, this task can be accomplished most efficiently by directing visual attention to the relevant entities as they are mentioned. In short, the listeners' gaze indicates the focus of their visual attention; where they direct their visual attention depends not only on the spoken utterances, but also on properties of objects or, more precisely, on the listeners' working memory representations of the objects (see Huettig et al., this issue), as well as higher-level inference processes (e.g., about the speaker's knowledge of the display), and the listeners' understanding of and compliance with the task demands. Similar comments apply, of course, to the production version of the paradigm. Here too the participants' eye gaze reflects their allocation of visual attention, and when and for how long the participants attend to the objects in a display depends not only on properties of the utterances they are preparing but also on properties of the display.

The fact that many variables can potentially influence the dependent measures can be seen as a strength of the paradigm. This is because in different versions of the paradigm all of these influences and their interactions can be studied. As illustrated in this review, the visual word paradigm has been used to study the way listeners understand and speakers produce utterances; it can also be used to study the processing of the objects in the display (e.g., the speed of the activation of their names), or to assess the performance of listeners or speakers who may have difficulty keeping the object representations in working memory or focusing on the task (e.g., Friedman-Hill, Robertson, & Treisman, 1995; Nation et al., 2003). Those versions of the paradigm where pairs of participants interact and instruct each other to move objects or request information about objects are excellent approximations to the way language is used in everyday conversational contexts (see also Tanenhaus & Trueswell, 2006).

An obvious limitation of the paradigm is that the speech that is presented or elicited always needs to be related to relevant visual input. For the production research, this means that the paradigm can only be used to study how speakers produce utterances about things they see (or, in a blank screen version of the production paradigm, things they have just seen). It is not obvious that the paradigm could be used to study how speakers generate utterances about past or future events, or how they verbalize abstract thoughts or emotions. The visual world paradigm can provide information about message generation and the formulation of utterances about the visual world, but we do not know how similar these processes are to those occurring when speakers do not talk about the visual world but their thoughts and feelings. One might, for instance, speculate that lexical access is supported by visual input and therefore faster in visual world experiments than in most other situations, or that linearization processes (deciding in which order to refer to different objects) are governed by the visual array and occur in a more orderly fashion than they normally do.

Because of the complexity of visual world experiments – the fact that both the production or comprehension of speech and the processing of a visual input are involved – the interpretation of the results is rarely straightforward. In comprehension studies in particular, it is often difficult to determine the contributions of visual and auditory processing to a pattern of findings. Critical issues are which representations of the visual and the auditory stimuli the listeners generate, and at which levels, and how the presence of the visual information affects the processing of the auditory information. This is important because most investigators do not aim to explain the participants' behavior in their specific visual world experiment; rather, they aim to generalize to a broader range of situations. Below, we discuss why such generalizations can be problematic (cf. Mitchell, 2004).

In visual world experiments, processing is based on both visual and the auditory input. Little is known about the properties of the listeners' representation of the speech input, and it is difficult to say how they might differ from representations they build up when they are presented with auditory information alone (see also Kamide, Altmann et al., 2003; Kamide, Scheepers et al., 2003). For instance, the spatial displays may discourage listeners from elaborative processes that would otherwise take place, or they may invite inferences that would not normally be drawn. The visual world studies by Altmann and colleagues have established that listeners hearing sentences such as “The boy will eat …” expect the continuation “the cake”, if this is the only edible object in the display, and that listeners hearing “The girl will ride the…” expect the continuation to be “carousel” rather than “motorbike”. These studies have demonstrated which kinds of information and knowledge listeners *can* draw upon when they process spoken utterances, but, as Altmann and colleagues have also pointed out, they do not imply that listeners will *always* draw upon these sources when they hear utterances. In fact, some of the most informative outcomes of visual world studies are perhaps those demonstrating that listeners do *not use*[2] certain types of information to direct eye gaze, e.g., when participants do not use stored object color knowledge to direct attention to objects shown in black and white (Huettig & Altmann, 2011).

The presence of a set of pictorial alternatives may also affect the way individual words are processed. Henderson and Ferreira (2004) have argued that the visual context may “lead to a situation in which the linguistic input is compared directly to the limited set of possibilities, rather than the natural case in which the input must generate the possibilities as well as selecting among them” (p. 48).

[2] The emphasis is on ‘use’; absence of a shift in eye gaze does not conclusively rule out activation of the relevant representation.

They proposed that time-course estimates about spoken word recognition based on visual world results may be biased towards greater speed. For instance, if listeners already shift their eye gaze towards the picture of a dog during the first moments of hearing "do..." this does not necessarily mean that people in general recognize spoken words that quickly. It may merely mean that little acoustic information is necessary to discriminate between the small number of alternatives present in the display "and that the system in essence bypasses normal acoustic lexical linguistic analysis and instead taps into cues that are chosen specifically to optimize performance in this task" (p. 48). However, Dahan and Tanenhaus (2004) have argued against this proposal, showing, for instance, intact target frequency effects (see also Dahan et al., 2001) and effects of neighborhood density (see also Magnuson et al., 2003; Magnuson et al., 2007), even when no competitor objects were displayed, which implies that listeners access the stored representations of the spoken words in their mental lexicon. Thus bypassing of lexical analyses (if this is possible at all) is unlikely.

However, the visual information may affect linguistic processing in more subtle ways. Most obviously, the speed of spoken word recognition may be affected through priming originating from the visual representations. As discussed previously, there is strong evidence that pictures of common objects rapidly activated their names (e.g., Huettig & McQueen, 2007; Meyer & Damian, 2007; Morsella & Miozzo, 2002). This can support the recognition of the spoken words in different ways. It could make the recognition of the words faster because their lexical representations are pre-activated. Recent data from a cross-modal priming study using the lexical decision task support this view. McQueen and Huettig (2005) found phonological inhibition (for picture primes of which the phonological onset was related to acoustically presented target pairs) and semantic facilitation (for same semantic category prime-target pairs), supporting the notion that the recognition of the spoken words can be affected by priming origination from viewing visual objects.

In addition, the speed of spoken word recognition may be affected because priming by vision-derived representations reduces ambiguity. There is a large number of words which have distinct meanings (e.g., "pen" referring to a writing implement or a cage) or senses (e.g., "chicken" referring to a whole animal or the meat) but identical phonology. Priming by a particular visual referent may speed up access of that meaning or sense on hearing the spoken word. In short, the visual world paradigm cannot be used to derive reliable estimates of the absolute time listeners need to recognize spoken words.

An additional type of interpretative problem may arise in studies where related competitors are presented in order to examine which alternatives become activated in the listeners' minds when they process a target. For instance, Allopenna et al. (1998, see Section 4.1.) showed that listeners hearing "beaker" looked both at the picture of an onset-related competitor (beetle) and a competitor with a rhyming name (speaker). It is unlikely that these results reflect strategic processes that participants develop de novo during the experiment; more likely listeners also activate word-initial and rhyme competitors in other contexts. Yet, it is difficult to gauge the importance of the weaker effects (the rhyme effect in the example). In the visual word paradigm, the representations of the competitor objects are held in (working) memory. This may lead to the activation of the corresponding lexical representations and render competitors more potent than they would be in other situations.

The opposite scenario can also occur: the results of a visual world experiment might lead one to underestimate the likelihood of a lexical candidate to become activated. This possibility was illustrated previously using the study by Dahan and Tanenhaus (2004) as an example but also applies to many other studies. Phonological competitors (e.g., "bot") of the target (e.g., "bok") were not fixated more than unrelated distractors when the target followed a semantically constraining context. However, this does not necessarily mean that the lexical representations of the competitor were not activated at all in the constraining context; instead it could mean that the activation of the competitor was not reflected in the eye gaze. This is because overt attention may reflect the outcome of a process which integrates context with initially exhaustive (i.e. form-based) lexical access. As already pointed out, eye gaze is a measure of overt attention and not a direct measure of the activation of lexical representations.

Finally, researchers may be interested in determining when or in which order different types of information associated with a word (e.g., the color of the referent object and semantic category information) become available. To assess this, one might use displays including a target and appropriate competitors. However, the interpretation of the results of such a study can be complicated by the fact that the timing of the listeners' eye movements depends on the number and properties of *all* objects in the display. For instance, Sorensen and Bailey (2007) found that the total number of displayed objects on a trial influenced the timing of shifts in eye gaze to semantic competitors. They observed that the larger the array size (four, nine, or sixteen objects), the later the competition effects occurred (see Huettig et al., this issue, for an account of working memory capacity limits of this finding). Huettig and Altmann (2011; see Section 4.4) found that when both a surface color competitor and a semantic competitor were present in the same visual display, participants directed their eye gaze 400 ms later to the color competitor than when the color competitor was the only competitor in the display even though the spoken sentences were identical across conditions. These findings illustrate that it is difficult to use the timing of shifts of eye gaze to competitors in a visual world experiment to estimate when particular kinds of information associated with a spoken word may become available in other circumstances.

In sum, the visual world paradigm is very well suited to studying how people produce and understand utterances about objects and events they see. This captures many everyday situations where people give or receive directions or instructions for action or talk about, say, the state of the kitchen floor, an abstract painting, or a reckless driver. Many of the cognitive processes occurring under such circumstances can be assessed using the visual world paradigm. It is, in particular, an excellent method for studying the interplay of language, vision, memory, and attention — cognitive processes that have traditionally been investigated in isolation, but that are of course all involved when language is used. Although one might often only be interested in one particular aspect of the language users' performance (e.g., their ability to discriminate different types of vowels), the empirical research should always be guided by a comprehensive theoretical model that encompasses all of the cognitive components involved in the task. This is because such a framework is necessary to estimate which general conclusions can be drawn from a specific set of findings.

Acknowledgements

We thank three anonymous reviewers for their comments on a previous version of this paper.

References

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.

Allport, D. A. (1971). Parallel encoding within and between elementary stimulus dimensions. *Perception & Psychophysics*, *10*, 104—108.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*, 79—87.

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. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502—518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye-movements and mental representation. *Cognition*, *111*, 55—71.
Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583—609.
Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163—187.
Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*, 63—80.
Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes*, *31*(2), 137—162.
Arnold, J. E., Altmann, R., Fagnano, M., & Tanenhaus, M. K. (2004). The old and thee, uh, new. *Psychological Science*, *15*, 578—582.
Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eyetracking evidence of the time-course of pronoun resolution. *Cognition*, *76*, B13—B26.
Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*, 25—36.
Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. K. (2007). If you say thee uh — You're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 914—930.
Arnold, J. E., Wasow, T., Losongco, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*, 28—55.
Bailey, K. G. B., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. 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. 485—500). Oxford, UK: Elsevier Inc.
Barr, D. J. (2008a). Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457—474.
Barr, D. J. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*, 18—40.
Barr, D. J., & Keysar, B. (2006). Perspective-taking and the coordination of meaning in language use. In M. J. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 901—938). (Second Edition). Amsterdam: Elsevier.
Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*, 441—455.
Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy aging. *Language and Cognitive Processes*, *22*, 1178—1210.
Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: What we can learn from free-viewing eye tracking* (pp. 249—278). New York: Psychology Press.
Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*, 653—658.
Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*, 896—904.
Brouwer, S. (2010). Processing strongly reduced forms in casual speech. *MPI Series in Psycholinguistics*, *57*, Wageningen: Ponsen & Looijen.
Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441—472.
Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*, 171—190.
Brown-Schmidt, S. (2009b). The role of executive function in perspective-taking during on-line language comprehension. *Psychonomic Bulletin and Review*, *16*, 893—900.
Brown-Schmidt, S. B., Byron, D., & Tanenhaus, M. K. (2004). That's not it and its not that: The role of conceptual composites in in-line reference resolution. In M. Carreiras & C. Clifton Jr. (Eds.), *On-line sentence processing: ERPS, eye movements and beyond* (pp. 2009—2228). : Psychology Press.
Brown-Schmidt, S. B., Byron, D., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*, 292—313.
Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122—1134.
Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*, 274—280.
Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*, 592—609.
Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643—684.
Canseco-Gonzalez, E., Brehm, L., Brick, C., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or Cárcel: Effects of age of acquisition and language proficiency on bilingual lexical access. *Language and Cognitive Processes*, *25*, 669—705.
Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention beyond a spotlight model. *Psychonomic Bulletin & Review*, *6*, 204—223.
Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*, 26—50.
Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains in real-time sentence comprehension. *Journal of Memory and Language*, *47*, 30—49.
Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Action-based affordances and syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *30*, 687—696.
Chomsky, N. (1957). Syntactic structures. Mouton: The Hague.
Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.
Clifton, C., Frazier, L., & Rayner, K. (Eds.). (1994). *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
Clifton, C., Jr., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, *2*, 234—250.
Clifton, C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317—334.
Cohen, J. D., Aston-Jones, G., & Gilzenrat, M. S. (2004). A systems-level perspective on attention and cognitive control: Guided activation, adaptive gating, conflict monitoring, and exploitation vs. exploration. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 71—90). New York: Guilford Press.
Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193—210.
Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886—899.
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*, 84—107.
Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*, 163—201.
Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*, 269—284.
Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317—367.
Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 498—513.
Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453—459.
Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292—314.
Davidoff, J., Davies, I., & Roberson, D. (1999). Color categories of a stone-age tribe. *Nature*, *398*, 203—204.
Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 237—262). New York: Plenum Press.
Dunabeitia, J. A., Aviles, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, *110*, 284—292.
Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*, 409—436.
Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*, 554—573.
Farmer, T., Anderson, S., & Spivey, M. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*, 570—595.
Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*, 405—410.
Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187—1199.
Fodor, J. (1983). *The modularity of mind*. Mass: MIT Press Cambridge.
Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030—1044.
Folk, C. L., Remington, R. W., & Johnston, J. C. (1993). Contingent attentional capture: A reply to Yantis (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 682—685.
Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimensional-weighting" account. *Perception & Psychophysics*, *58*, 88—101.
Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. Ph.D. Dissertation. Indiana University Linguistics Club. University of Connecticut.
Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559—586). Hillsdale, NJ: Lawrence Erlbaum Associates.
Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, *24*, 437—468.
Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, *269*, 853—855.
Frost, R. (1998). Towards a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*, 71—99.

Gennari, S. P., Sloman, S., Malt, B., & Fitch, T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49—79.
Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*, 489—494.
Gleitman, L., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544—569.
Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23—45.
Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts (Vol. III)* (pp. 41—58).
Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1—B14.
Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review*, *10*(3), 603—609.
Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: What we can learn from free-viewing eye tracking* (pp. 213—247). New York: Psychology Press.
Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274—279.
Griffin, Z. M., & Oppenheimer, D. M. (2006a). Looking and lying: Speakers' gazes. Reflect locus of attention, not content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 943—948.
Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42—55.
Grodner, D. J., & Sedivy, J. (in press). The effects of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (eds). *The processing and acquisition of reference.* MIT Press: Cambridge, MA.
Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, *3*(3), 233—274.
Grosjean, F., & Frauenfelder, U. H. (1996). Spoken word recognition paradigms. *Special Issue of Language and Cognitive Processes*, *11*(6).
Gumperz, J. J., & Levinson, S. C. (1996). Rethinking linguistic relativity. Cambridge: Cambridge University Press.
Hagoort, P., Hald, L. A., Bastiaansen, M. C. M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304* (5669), 438—441.
Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105—115.
Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43—61.
Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113—157.
Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech*, *14*, 277—288.
Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of references. *Cognition*, *108*, 811—836.
Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action* (pp. 1—58). Hove: Psychology Press.
Howell, D. C. (2002). Statistical methods for psychology. Pacific Grove, CA: Duxbury.
Huang, Y., & Snedeker, J. (2005). What exactly do numbers mean? *Paper presented at the Experimental Pragmatics Conference*. UK: Cambridge.
Huang, Y., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376—415.
Huang, Y., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in five-year olds: Evidence from real time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723—1739.
Huang, Y. & Snedeker, J. (in press). Logic and Conversation revisited: Evidence for the division between semantic and pragmatic content in real time language comprehension. *Language and Cognitive Processes.*
Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 187—207). New York, NY: Psychology Press.
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., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*, 985—1018.
Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing 'frog' – How object surface color and stored object color knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology*, *64*, 122—145.
Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, *10*, 39—58.
Huettig, F., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition*, *36*, 341—360.
Huettig, F., & Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes*, *25*, 347—374.
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*, *54*, 460—482.
Huettig, F., & McQueen, J. M. (2008). Retrieval and use of components of lexical knowledge depend on situational demands. *Paper presented at the AMLaP 2008 conference in Cambridge, UK*.
Huettig, F., & McQueen, J. M. (2009). AM radio noise changes the dynamics of spoken word recognition. *Paper presented at the AMLaP 2009 conference in Barcelona, Spain*.
Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, *121*, 65—80.
Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral & Brain Sciences*, *24*, 453—509.
Humphreys, G. W., Riddoch, M. J., & Price, C. J. (1997). Top-down processes in object identification: Evidence from experimental psychology, neuropsychology, and functional anatomy. *Philosophical Transactions: Biological Sciences*, *352*, 1275—1282.
Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 105—134). New York: Psychology Press.
Ito, K., & Speer, S. R. (2008). Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*, 541—573.
Johnson, E. K., & Huettig, F. (2011). Eye movements during language-mediated visual search reveal a strong link between overt visual attention and lexical processing in 36-month-olds. *Psychological Research*, *75*, 35—42.
Johnson, E. K., Huettig, F., & McQueen, J. M. (2008). Conceptual attributes of heard words modulate toddlers' attention to the visual scene. *Paper presented at the 11th International Congress for the Study of Child Language (IASCL), Edinburgh, Scotland, UK*.
Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, *15*, 314—318.
Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*, 55—80.
Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113—147.
Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*, 709—748.
Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647—670.
Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133—156.
Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37—55.
Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32—38.
Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25—41.
Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*, 481—529.
Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye-movements. *Journal of Memory and Language*, *57*, 519—543.
Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*, 95—127.
Kumada, T. (2001). Feature-based control of attention: Evidence for two forms of dimension weighting. *Perception & Psychophysics*, *63*, 698—708.
Lackner, J. R., & Tuller, B. H. (1979). Roles of efference monitoring in the detection of selfproduced speech errors. In W. E. Cooper & E.C.T. Walker (Eds.), *Sentence processing. Psycholinguistic studies presented to Merrill Garrett* (pp. 281—294). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Land, M. F., Mennie, N., & Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception*, *28*, 1311—1328.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211—240.
Laver, J. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.
Levelt, W. J. M. (1989). Speaking. From intention to articulation. Cambridge, MA: MIT Press.
Levelt, W. J. M., & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, *12*(4), 433—452.
Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language*, *35*(6), 757—774.
Lockridge, C. B., & Brennan, S. E. (2001). Addressees' needs affect speakers' syntactic choices. *11th Annual meeting of the society for text and discourse, UC Santa Barbara*.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, *39*, 367—386.
Lucy, J. A. (1992). Language diversity and thought: A reformulation of the linguistic relativity hypothesis. Cambridge: Cambridge University Press.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676—703.
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.
Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, *108*, 866—873.
Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*, 202—227.
Malpass, D. & Meyer, A. S. (2010). Parallel processing of objects in a naming task: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*, 173—193.
Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between language competition. *Bilingualism: Language and Cognition*, *6*, 97—115.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken wordrecognition. *Cognition*, *25*, 71—102.
Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistics and computational perspectives* (pp. 148—172). Cambridge, MA: MIT Press.
Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29—63.
Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: On the importance of word onset. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576—585.
Mayberry, M., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of the coordinated interplay of utterance, visual context, and world knowledge. *Cognitive Science*, *33*, 449—496.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1—86.
McDonald, S. A. (2000). *Environmental determinants of lexical processing effort*. Unpublished doctoral dissertation, University of Edinburgh, Scotland. Retrieved December 10, 2004, from http://www.inf.ed.ac.uk/publications/thesis/online/IP000007.pdf
McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., & Subik, D. (2008). Gradient sensitivity to within-category variation in speech: Implications for categorical perception. *Journal of Experimental Psychology, Human Perception and Performance*, *34*, 1609—1631.
McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*, 1—39.
McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33—B42.
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*, 65—91.
McQueen, J. M., Dahan, D., & Cutler, A. (2003). Continuity and gradedness in speech processing. In N. O. Schiller & A.S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 39—78). Berlin: Mouton de Gruyter.
McQueen, J. M., & Huettig, F. (2005). Semantic and phonological priming of auditory lexical decision by pictures and printed words. *Paper presented at the AMLaP 2005 conference in Ghent, Belgium*.
McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*, 661—671.
McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in online sentence comprehension. *Journal of Memory and Language*, *38*, 283—312.
Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14*, 710—716.
Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture–picture interference paradigm. *Memory & Cognition*, *35*, 494—503.
Meyer, A. S., Ouellet, M., & Häcker, C. (2008). Parallel processing of objects in a naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 982—987.
Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*(1), 131—147.
Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25—B33.
Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*, 314—319.
Meyer, A. S., van der Meulen, F., & Brooks, A. (2004). Eye movements during speech planning: Speaking about present and remembered objects. *Visual Cognition*, *11*, 553—576.
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.
Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, *37*, 1026—1039.
Mishra, R. K., & Singh, N. (2010). Online fictive motion understanding: An eye-movement study with Hindi. *Metaphor & Symbol*, *25*, 144—161.
Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras, & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 15—32). New York, NY: Psychology Press.
Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 244—263.
Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin*, *132*, 297—326.
Morgan, J., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 428—442.
Morgan, J. L., van Elswijk, G., & Meyer, A. S. (2008). The time-course of the phonological activation of successive object names: Evidence from word probe experiments. *Psychonomic Bulletin & Review*, *15*, 561—565.
Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 555—563.
Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to word meanings during spoken language comprehension: Effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 1254—1276.
Moss, H. E., McCormick, S. F., & Tyler, L. K. (1997). The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes*, *12*, 695—731.
Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*, 578—594.
Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, *57*, 1—17.
Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*, 329—336.
Nation, K., Marshall, C. M., & Altmann, G. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*, 314—329.
Navarette, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*, 359—377.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189—234.
Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357—395.
Onifer, W., & Swinney, D. (1981). Accessing lexical ambiguity during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition*, *9*, 225—236.
Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*, 528—537.
Pallier, C., Colome, A., & Sebastian-Galles, N. (2001). The influence of nativelanguage phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, *12*(6), 445—449.
Panizza, D., Chierchia, G., Huang, Y., & Snedeker, J. (in press). The Relevance of polarity for the online interpretation of scalar terms. The Proceedings of Semantics and Linguistic Theory (SALT) 19.
Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 155—184.
Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*, 189—219.
Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, *98*, B75—B87.
Papafragou, A., & Trueswell, J. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, *63*, 64—82.
Postma, A. (2000). Detection of errors during speech production. A review of speech monitoring models. *Cognition*, *77*, 97—131.
Pratt, J., & Hommel, B. (2003). Symbolic control of visual attention: The role of working memory and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 835—845.
Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately. *Quarterly Journal of Experimental Psychology*, *63*(4), 772—783.
Richardson, D. C., Altmann, G. T. M., Spivey, M. J., & Hoover, M. A. (2009). Much ado about eye movements to nothing. *Trends in Cognitive Sciences*, *13*, 235—236.
Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269—295.
Roberson, D., Davies, I., & Davidoff, J. (2000). Colour categories are not universal: Replications and new evidence from a Stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369—398.
Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107—142.
Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language*, *57*, 232—251.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353—368.
Roy, D., & Mukherjee, N. (2005). Towards situated speech understanding: Visual context priming of language models. *Computer Speech and Language*, *19*, 227—248.
Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1108—1117.
Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51—89.
Sapir, E. (1921). Language: An introduction to the study of speech. New York, NY: Harcourt, Brace and Company.
Scheepers, C. (2003). Syntactic priming of relative clause attachment: Persistence of structural configuration in sentence production. *Cognition*, *89*, 179—205.
Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123—164). Hillsdale, NJ: Lawrence Erlbaum.
Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211—232.
Sedivy, J., Tanenhaus, M. K., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109—147.
Seidenberg, M., Tanenhaus, M., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of the knowledge-based processing. *Cognitive Psychology*, *14*, 489—537.
Sereno, Cameron, & O'Donnell (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, *14*(4), 328—333.
Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, *17*, 372—377.
Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 24—49). Cambridge, MA: MIT Press.
Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. J. Gumperz & S.C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70—96). Cambridge: Cambridge University Press.
Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the investigation of language and thought* (pp. 157—191). Cambridge, MA: MIT Press.
Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*, 103—130.
Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238—299.
Sorensen, D. W., & Bailey, K. G. D. (2007). The world is too much: Effects of array size on the link between language comprehension and eye movements. *Visual Cognition*, *15*, 112—115.
Spivey, M., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235—241.
Spivey, M., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors: Thinking with your hands. *Proceedings of the National Academy of Sciences*, *102*, 10393—10398.
Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281—284.
Spivey, M. J., Richardson, D. C., & Fitneva, S. A. (2004). Memory outside of the brain: Oculomotor indexes to visual and linguistic information. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
Stephen, D. G., Mirman, D., Magnuson, J. S., & Dixon, J. A. (2009). Lévy-like diffusion in eye movements during spoken-language comprehension. *Physical Review, E*, *79*, 056114.
Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language*, *46*, 39—56.
Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 153—162.
Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, *49*, 161—167.
Tan, L. H., Chan, A. H. D., Kay, P., Khong, P. -L., Yip, L. K. C., & Luke, K. K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Science*, *105*, 4004—4009.
Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society, B*, *363*, 1105—1122.
Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous ds in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*, 427—440.
Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*, 557—580.
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.
Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 863—900). (second edition). New York: Elsevier Academic Press.
Thompson, C. K., & Choy, J. J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research*, *38*, 255—283.
Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, *40A*, 201—237.
Trueswell, J., & Gleitman, L. R. (2004). Children's eye movements during listening: Evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (Eds.), *Interface of language, vision, and action: Eye movements and the visual world* (pp. 319—346). NY: Psychology Press.
Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89—134.
Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 155—179). Hillsdale, NJ: Lawrence Erlbaum Assoc.
Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: The use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285—318.
Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning of Verbal Behavior*, *16*, 683—692.
Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701—718.
Van der Wel, R. P. R. D., Eder, J. R., Mitchel, A. D., Walsh, M. W., & Rosenbaum, D. A. (2005). Trajectories emerging from discrete versus continuous processing models in phonological competitor tasks: A commentary on Spivey. In Grosjean & Knoblich (Eds.), *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 588—594.
Van Gompel, R. P. G., & Pickering, M. J. (2001). Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998). *Psychonomic Bulletin and Review*, *8*, 851—857.
Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284—307.
Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 371—386.
Walsh, M., Dickey, J. J., Choy, J. W., & Thompson, C. K. (2007). Real-time comprehension of wh- movement in aphasia: Evidence from eyetracking while listening. *Brain and Language*, *100*, 1—22.
Watson, J. B. (1925). Behaviorism. New York: Norton.
Watson, D. G., Tanenhaus, M., & Gunlogson, C. (2008). Interpreting pitch accents in on-line comprehension: H* vs LH*. *Cognitive Science*, *32*, 1232—1244.
Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*, 367—392.
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1—25.
Weber, A., Melinger, A., & Lara Tapia, L. (2007). The mapping of phonetic information to lexical representation in Spanish: Evidence from eye movements. In J. Trouvain & W.J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1941—1944). Dudweiler: Pirrot.
Wheeldon, L., Meyer, A. S., & van der Meulen, F. F. (2007). Speech to gaze alignment in anticipation errors. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye-movements: A window on mind and brain*. Oxford: Elsevier.
Whitney, P., McKay, T., Kellas, G., & Emerson, W. A. (1985). Semantic activation of noun concepts in context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 126—135.
Whorf, B. J. (1956). Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. In John B. Carroll (Ed.), Cambridge: MIT Press.
Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). The Russian Blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science*, *104*, 7780—7785.
Wühr, P., & Frings, C. (2008). A case for inhibition: Visual attention suppresses the processing of irrelevant objects. *Journal of Experimental Psychology: General*, *137*, 116—130.
Wühr, P., & Waszak, F. (2003). Object-based attentional selection can modulate the Stroop Effect. *Memory & Cognition*, *31*(6), 983—994.
Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of Cognitive Neuroscience*, *20*, 592—612.
Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2009). Function follows form: Activation of shape & function features during concept retrieval. *Poster presented at the 50th Annual Meeting of the Psychonomic Society*.
Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin and Review*, *16*(5), 869—874.
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—14.