Language and Linguistics Compass 10/2 (2016): 66–82, 10.1111/lnc3.12177

Visually Situated Language Comprehension

Pia Knoeferle[1]* and Ernesto Guerra[2]

[1] Department of German Language and Linguistics, Humboldt University, Berlin

[2] Experimental Psycholinguistics Lab and Interdisciplinary Center for Intercultural and Indigenous Studies, Pontificia Universidad Católica de Chile

Abstract

Over the past two decades, 'visually situated' language comprehension (the interplay between language comprehension, attention, and non-linguistic visual context) has emerged as an increasingly active area of research. One important result in this area is that both linguistic and world knowledge, as well as visual cues, can rapidly inform the unfolding interpretation as reflected by comprehenders' eye movements to objects during spoken language comprehension. However, upon closer inspection, temporal delays of object-directed gaze are not infrequent and emerge for the processing of non-canonical (vs. canonical) structures, for scalar implicatures and for recently learned world–language associations. While it may further be tempting to assume that the different knowledge sources and visual cues are on a par in guiding visual attention, comprehenders' eye movements in many instances reveal a robust referential priority (more looks go to the referent of a word than to other objects). Should this priority be taken as a trivial observation? In the present article, we argue that the tension between this referential priority and other world–language relations constitutes an important constraint on the linking hypotheses and mechanisms implicated in situated language comprehension and should be considered when conceptualizing models and accounts of visually situated language comprehension.

1. *Introduction*

The question of how language performance is related to, and how it benefits from, information in the non-linguistic visual context (henceforth: visual context) has come up time and again in scientific research in different forms. Up to approximately 1980, researchers examined among other topics the effects of pictorial information on language acquisition and learning (e.g., Deno 1968; Kellogg and Howe 1971; Moeser and Bregman 1972; Moeser and Olson 1974), memory for words, sentences, and pictures (e.g., Pezdek 1977; Shepard 1967), the mechanism implicated in verifying a sentence against a picture (e.g., Clark and Chase 1972; Gough 1966; Just and Carpenter 1971; Tanenhaus, Carroll, and Bever 1976), and the mental representations of linguistic and pictorial information (e.g., Paivio 1971; Potter and Faulconer 1975). The 1980s and early 1990s saw a continued investigation of these topics (language acquisition and learning: e.g., Whitehurst, Arnold, Epstein, Angell, Smith, and Fischel 1988, Whitehurst, Falco, Lonigan, Fischel, DeBaryshe, Valdez-Menchaca, and Caulfield 1994; Tomasello and Farrar 1986; memory: e.g., Simons 1996; picture–sentence verification: e.g., Kroll and Corrigan 1981; Marquer and Pereira 1990; Mathews, Hunt, and MacLeod 1980; and linguistic and pictorial mental representations: e.g., Paivio 1986; Potter, Kroll, Yachzel, Carpenter, and Sherman 1986). One conclusion from this research is that visual context can modulate a broad range of cognitive processes, supporting a view of language and cognition in which the immediate visual context plays an important role. The precise time course of how visual cues are integrated during language processing, however, remained elusive prior to the 1990s (but see Cooper 1974).

From around 1990, researchers began to examine the time course of language and picture processing using event-related brain potentials (ERPs, electrical brain activity recorded at the scalp, time-locked to the presentation of a stimulus). In the ERP paradigms of the 1990s, participants were typically seated in front of a computer display and inspected stimuli (e.g., written words or pictures) presented in rapid serial visual presentation. The recording of ERPs in this type of paradigm has revealed rapid deviations in the electrical brain activity as a function of the semantic fit of a stimulus in context. These deviations manifest themselves as increased mean amplitude negativities peaking around 400 ms (N400). The N400 was first discovered for language by Kutas and Hillyard (1984), but negativities also emerged for picture processing (e.g., Barrett and Rugg 1990; Nigam, Hoffman, and Simons 1992) and when processing pictures in a linguistic context (Ganis, Kutas, and Sereno 1996), although with subtly differing topographies reflecting sensitivity to stimulus modality (see Kutas and Federmeier 2011).

Following a publication by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995),[1] which introduced eye movements during spoken comprehension as a new measure, psycholinguists and cognitive scientists rapidly adopted the so-called 'visual-world' paradigm. In this paradigm, participants are seated either in front of a computer display, showing a scene, or in front of real-world objects. An eye tracker records a participant's gaze to these objects moment to moment (in millisecond resolution) as she listens to a sentence and performs a task (e.g., passive listening, responding to comprehension questions, picture–sentence verification, or object manipulation to name some of the more common tasks; see Figure 1 for an example eye tracker setup). The software outputs both the *x*–*y* coordinates and time stamps of a participant's fixations, and we can link these data to object positions and to the onset times for visual and linguistic stimuli. In this way, we can relate spoken words in the unfolding utterance to object-directed fixations (e.g., Tanenhaus et al. 1995; see Huettig, Rommers, and Meyer 2011; Tanenhaus and Trueswell 2006 for reviews). ERP versions of this paradigm do also exist (see Knoeferle 2015a, for a review), but we will focus on eye-tracking results for the present article.

What insights into language comprehension could we hope to gain from eye movements to objects during spoken language comprehension? Certainly, they have revealed how rapidly different kinds of information (our linguistic and world knowledge and reference to the visual context) guide our visual attention, and, by association, they have provided insight into the implicated cognitive mechanisms (e.g., a tendency to anticipate upcoming information;

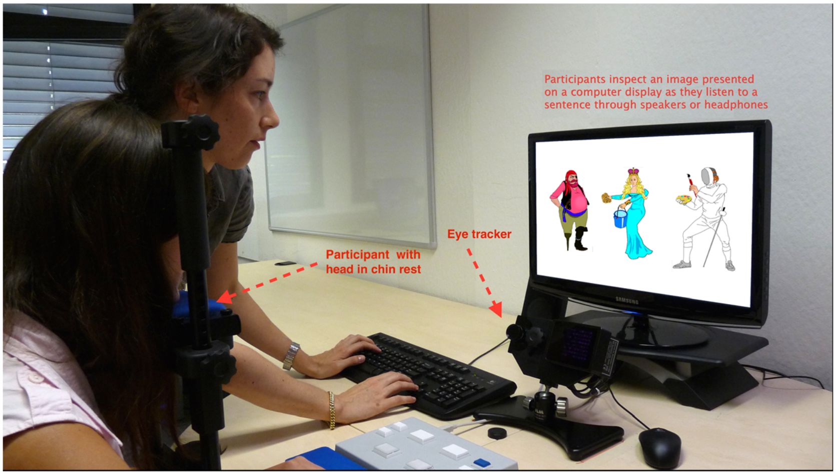
Participants inspect an image presented on a computer display as they listen to a sentence through speakers or headphones
Eye tracker
Participant with head in chin rest

Fig 1. Example eye tracker setup. The participant is seated in front of a display with her head in a chin rest. An eye tracker (SR Research), located at the foot of the monitor and facing the participant, tracks a participant's eye movements to objects on the display. The eye tracker software outputs the *x*–*y* coordinates and computes the exact time stamps of these fixations relative to the presentation of a picture and a spoken sentence.

Kamide 2008). In this context, existing evidence shows that all sorts of world–language relations (e.g., referential, lexico-semantic, and compositional) can be interpreted rapidly and incrementally. However, the temporal coordination of utterance interpretation and object-directed visual attention is not invariantly rapid. Relative delays in the time course emerge, perhaps unsurprisingly, as a function of (non-canonical) word order (e.g., Kamide, Scheepers, and Altmann 2003a, Knoeferle, Crocker, Scheepers, and Pickering 2005, Weber, Grice, and Crocker 2006), newly learned (vs. long-term) world–language associations (e.g., Creel, Aslin, and Tanenhaus 2008), and complex pragmatic inferences (e.g., Huang and Snedeker 2009; Section 2.2). These observations suggest that the close time-locking of gaze to utterance interpretation varies with the situation and implicated mental processes.

In light of such variation in visual attention, one may ask what systematicity, if any, underlies the eye-movement record. To gain insight into potential systematicity, extant research has asked whether different cues or world–language relations are on a par in informing language processing. Visually situated approaches have for instance argued that all else being equal, comprehenders exhibit a 'referential priority / preference' (see Knoeferle and Crocker 2006, 2007; Knoeferle, Carminati, Abashidze, and Essig 2011). By 'referential', we mean the relation between a word and the object it denotes, and this includes at least noun–object and verb–action relations (Jackendoff 2002). For nouns, a referential priority manifests itself in more looks to an object that is named than objects related to language in another manner (e.g., through lexical associations). Referential priority also captures the observation that comprehenders prefer to relate a sentential verb to a depicted action (and its associated agent) over relating it to a stereotypically associated agent, arguably reflecting a priority in interpretation. The preference has mostly been assessed in terms of an increased gaze probability though some evidence suggests that referential looks also occur earlier (see Scheepers, Keller, and Lapata 2008; see Knoeferle and Crocker 2006 for a relevant processing account).

This preference is not absolute, and of course, other world–language relations also modulate comprehenders' visual attention but often less so than referential relations. For instance, upon hearing that somebody is spied upon, comprehenders looking for a possible agent more frequently inspect a wizard depicted as spying than a detective (stereotypically associated with the spying, Knoeferle and Crocker 2006). The priority also captures the finding that upon hearing a subject–verb sentence beginning (e.g., 'The waiter polish…'), more attention goes to the location at which the polishing action had just been depicted and at which the action target is located (e.g., candelabra that the waiter polished) than to the (different) location and target of a future polishing action (e.g., crystal glasses, Knoeferle and Crocker 2007, Experiment 3; henceforth 'recent-event preference').

One level of analysis that one might want to consider in accommodating this preference is probabilistic. Indeed, the importance of probabilistic information has shaped many accounts of language processing (e.g., constraint-based lexicalist accounts: MacDonald, Pearlmutter, and Seidenberg 1994; Trueswell and Tanenhaus, 1994; probabilistic information-theoretic approaches, e.g., Hale 2001; Levy 2008) and has also been considered in research on visually situated language processing. For instance, it has been argued that all cues are equally important, except that one may be more 'predictive' than the other regarding the upcoming input (e.g., Altmann and Mirkovic 2009, p. 586). Predictiveness, consequently, could be defined in terms of the frequency of a cue and cue frequency and in turn could accommodate a referential priority. However, corpus analyses revealed that at least in the case of verb-mediated recent events, this preference is not caused by the long-term frequency of linguistic expressions (Knoeferle et al. 2011), nor is it eliminated when pitted against either a very strong short-term frequency bias (how often people see a future vs. recent event and hear it mentioned) or situation-immediate cues such as an actor's gaze (Abashidze, Knoeferle, and Carminati 2014, 2015).

This is not to say that the preference is invariant; in fact, while opposing short-term frequency biases did not eliminate the preference, they modulated it, corroborating that probabilistic information plays an important role in language comprehension.

The present article, however, emphasizes a complementary psycholinguistic level of analysis – with an eye to what visual attention can reflect about different world–language relations, linking hypotheses, and associated comprehension processes. In the next section, we discuss evidence on the sensitivity of visual attention to different (e.g., phonological, lexico-semantic, syntactic, and pragmatic) comprehension processes and highlight situations and mental operations that have elicited a relative delay of object-based gaze.

At the same time, we note that comprehenders inspect referents more than objects linked through other (e.g., lexico-semantic) associations. These observations permit us to constrain accounts of situated language processing and to complement existing linking hypotheses.[2] The latter have highlighted either the link between fixating an object and the activation of its lexical representation (i.e., its name, Tanenhaus, Magnuson, Dahan, and Chambers 2000) or the link between conceptual representations from the linguistic input (be they lexical or compositional) and from the scene (Altmann and Kamide 2007). What this paper contributes is insight into the relative weighting of referential relative to non-referential linking hypotheses and their associated comprehension processes. In the third section, we accordingly review further results in support of a referential priority. In the last section, we argue that assessing the importance of information types and of different world–language relations can constrain the linking hypotheses and mechanisms implicated in situated language comprehension.

2. *Sensitivity of Visual Attention to Comprehension Processes*

In this section, we discuss evidence in the literature suggesting that visual attention is sensitive to referential relations but also subtler lexical associations (Section 2.1), to structural variation, and complex pragmatic inferences (Section 2.2). We will argue that many world–language relations are processed highly rapidly but that delays also emerge, for the processing of non-canonical structure, newly learned associations, and scalar implicature. We also note that visual attention is particularly sensitive to referential expressions, such that more looks go to an object when it is named than when it is otherwise related to language (e.g., through lexico-semantic associations).

2.1. OBJECT-DIRECTED VISUAL ATTENTION: LEXICAL PROCESSES

In situated language comprehension, the time from when listeners begin to process a word to when they shift their gaze to its referent has been taken to reflect processes of establishing reference. Other (unnamed) objects temporarily compete for attention (henceforth: ‘competitors’), for instance, because they resemble the referent in name or shape, or because they belong to the same conceptual category. The (temporary) deflection of visual attention to these other objects has been interpreted as indexing the activation of phonological (e.g., when overlap is in the name) or of lexico-semantic knowledge (e.g., when referent and competitor belong to the same conceptual category). For a word such as *beaker*, participants began to inspect both the picture of a beaker and a picture of a phonological competitor (a beetle) more often than unrelated targets from around 200 ms after word onset (e.g., Allopenna, Magnuson, and Tanenhaus 1998; see also Tanenhaus et al. 1995 and Dahan 2010).

The deflection of visual attention to the phonological competitor is typically short-lived (from 200 to 700 ms after the onset of *beaker*) and begins to decay shortly after the offset of *beaker* (around 400 ms after its onset). Even competitors such as a speaker – whose name rhymes with the target word *beaker* – attracted more looks from around 300 ms after the onset of the word *beaker* than phonologically unrelated objects (but see Ben-David Chambers, Daneman, Pichora-Fuller, Reingold, and Schneider 2011 for evidence on age-related changes of these processes and Yee, Blumstein, and Sedivy 2008 on evidence for aphasics; see Salverda, Dahan, and McQueen 2003 on the effects of words within other words; and see Altmann 2011 on the time course of language-mediated eye movements). Thus, reference is disambiguated within a few hundred milliseconds, a process during which phonological competitors are also activated and temporarily attract visual attention. After disambiguation, most attention goes to the referent while attention to the competitors begins to decrease.

Perhaps unsurprisingly, other (non-phonological) relations between a word and an object also re-direct some visual attention to a competitor. Among these relations are shared surface color, category membership, or shape. For instance, when listeners heard … *he looked at the piano*, their eye gaze shifted to a piano on a higher proportion of trials than to unrelated objects, but when no piano was visible, other unnamed but semantically related objects (a trumpet) attracted more attention between 200–300 and 800 ms than entirely unrelated objects. When both a piano and a trumpet were present, most looks went to the piano, but the trumpet attracted more looks than semantically unrelated objects (Huettig and Altmann 2005; see Yee and Sedivy 2006). Competitors were also fixated if they were depicted in the shape typical of the named object. When participants were instructed to move a snake (depicted as stretched out) to another location, a nearby competitor (a rope) depicted in a prototypical snake shape (coiled up) was inspected less often than the snake but more often than unrelated objects (Dahan and Tanenhaus 2005). These eye-movement differences occurred between approximately 200–300 and 1100 ms after target word onset.

Subtle variation in the time course of object-directed looks emerged when world–language associations were newly learned. Creel, Aslin, and Tanenhaus (2008, Experiment 1) examined the effects of a talker's voice (male vs. female) on the processing of lexical competitors. When listeners had heard the two words of a cohort pair pronounced repeatedly by different talkers (a male voice uttered *sheep*; a female voice *sheet*), they made fewer fixations to the competitor object than when the same speaker had pronounced both words of a pair. Analyses are reported for a 200- to 800-ms time window post-target word onset, and, descriptively, talker modulation of visual attention to the competitor object occurred between 500 and 600 ms post-word onset and lasted until 1100 ms. Such relative delay suggests that recently acquired associations take slightly more time to influence language comprehension than long-term linguistic and world knowledge.

In summary, as listeners encounter a word, they rapidly direct looks to the word's referent. In addition, some (but less) attention goes to other objects that overlap in name or that share other (conceptual and perceptual) features with the referenced object. When the associations between referents and the linguistic input (of a speaker's voice) were learned short term, talker modulation of visual attention to the competitor object occurred with some delay, suggesting that the time course of language-mediated visual attention can reflect differences in information encoding.[3] Unsurprisingly, referential expressions elicit more visual attention to an object than linguistic expressions that are otherwise related to that object (e.g., through lexico-semantic-associations), suggesting that referential relations are prioritized in the linking assumptions and comprehension mechanisms.

2.2. DELAYED GAZE SHIFTS: NON-CANONICAL ORDER AND SCALAR IMPLICATURE

While many lexical language–word relations rapidly affect object-directed visual attention (but see, e.g., Creel et al. 2008, Section 2.1), delays emerge when comprehenders process non-canonical sentence structures or complex pragmatic implicatures. With regard to sentence structure, for instance, subject–verb–object (SVO) sentences in German are canonical, while object–verb–subject (OVS) sentences are non-canonical; the latter take longer to read, suggesting that they are more difficult to process (e.g., Hemforth 1993; Knoeferle and Crocker 2009), and when they are initially structurally ambiguous, they elicit increased mean amplitude positivities[4] in ERPs, suggesting a revision of sentence structure (e.g., Matzke, Mai, Nager, Rüsseler, and Münte 2002). The difficulty associated with the processing of the OVS order is also reflected in visually situated language studies, as is its modulation by prosody or the visual context.

Weber et al. (2006), for instance, monitored participants' eye movements to objects as participants listened to German NP-V-NP sentences in which the first noun was ambiguous in grammatical function and thematic role (SVO: *Die Katze jagt wohlmöglich den Vogel*, 'The cat (amb.) chases possibly the bird (obj)'; or OVS: *Die Katze jagt wohlmöglich der Hund*, 'The cat (amb.) chases possibly the dog (subj))'. They manipulated prosody to cue word order (for SVO sentences, the nuclear accent was on the verb; for OVS sentences, it was on the first noun phrase). Participants initially anticipated a character that was a plausible sentential object (but not subject: the bird), even when the intonation supported the OVS structure; gaze pattern reflecting disambiguation emerged only after the verb. At this point, participants anticipated the agent of a cat-chasing event (the dog) more for OVS than for SVO prosody, suggesting that they had assigned a subject function and agent role to 'dog' and an object function and patient role to 'the cat'. Note that these looks are 'anticipatory', viz. participants inspect the dog more *before* it is named, and could thus be viewed as rapid responses. But relative to when participants have the information necessary to identify the dog as the correct role filler (at the verb), these effects are delayed by one word (post-verbal).

The gaze behavior observed by Weber et al. emerged post-verbally and thus subtly delayed relative to the compositional effects of noun meaning, verb meaning, and associated world knowledge, which began to emerge during the verb (Kamide, Altmann, and Haywood 2003b). It is possible that the delay results from difficulty associated with processing the non-canonical OVS structure, or, alternatively, from comparatively weaker effects of prosody as a cue to sentence structure (see Sedivy, Tanenhaus, Chambers, and Carlson 1999, Experiment 1B).

In a different study (Knoeferle et al. 2005), a sentence-final case-marked noun phrase disambiguated initially structurally ambiguous utterances towards either SVO or OVS. Early disambiguation was possible when the verb referred either to an action (washing) of the first-mentioned referent (a princess, SVO, e.g., *Die Prinzessin wäscht offensichtlich den Pirat*, 'The princess (amb, obj) washes apparently the pirate (subj)') or to another action that depicted the princess as the object and patient (OVS, e.g., *Die Prinzessin malt offensichtlich der Fechter*, 'The princess (amb, obj) paints apparently the fencer (subj)'; see Figure 2). During the verb, eye movements went more often to the pirate than the fencer for both SVO and OVS sentences. Post-verbally, however, before the fencer was mentioned, they also reflected participants' expectations of the OVS order and patient–agent role relations. For OVS (compared with SVO) sentences, participants gazed more often at the fencer, the subject, and agent of the other depicted event. Interestingly, these eye movements occurred with a time course reminiscent of the effects of prosody on structural disambiguation (Weber et al. 2006) and also reminiscent of the combined effects of case marking, verb meaning, and world knowledge in German SVO/OVS sentences (Kamide et al. 2003a, all post-verbal effects, but see Kaiser and Trueswell 2004 on relevant evidence that discourse-based expectations can eliminate this difficulty in another language that permits scrambling, viz. Finnish).

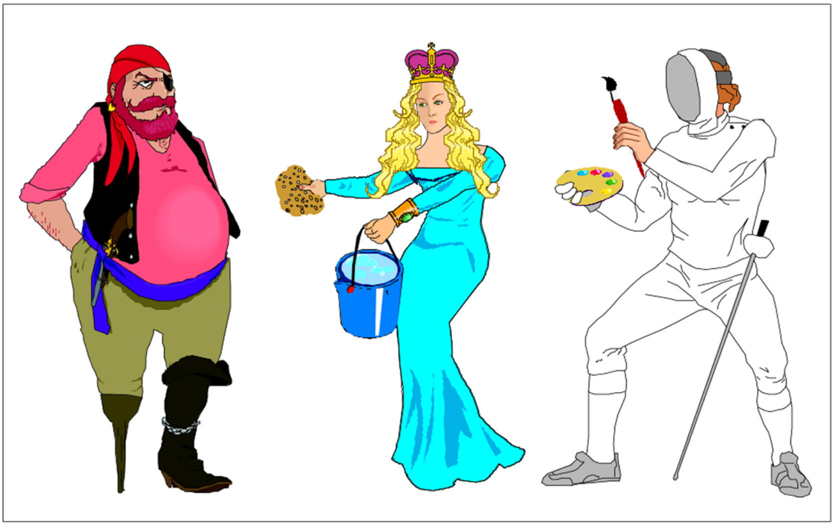

Fig 2. Example image from Knoeferle et al. (2005), p. 100.

Delayed effects on listeners' visual attention emerged also for the quantifier *some* when it involved computing scalar implicature. In a study by Huang and Snedeker (2009), a depicted girl and a boy were each 'given' two (depicted) socks, and another girl was given three balls. When participants heard *Point to the girl that has some/two of the* …, the girl with two socks was inspected substantially later for *some* (around 1000 ms after word onset) than for *two* (target inspection rose above chance in the first 200 ms after quantifier onset, Experiments 1 and 2). The authors attributed the delay to the computation of a scalar implicature since gaze pattern suggested that *some* was interpreted without delay when its meaning disambiguated reference (this was the case when nine socks were evenly distributed among two boys and one girl, while another girl had no socks, and participants heard *Point to the girl that has some of the*… Experiment 3; but see Grodner, Klein, Carbary, and Tanenhaus 2010).

By contrast, short-term talker associations (which seemed to affect object-directed attention with some delay for lexical ambiguity resolution) did not result in a delay of object-directed visual attention for the resolution of structural ambiguity. Kamide (2012) trained participants to associate one critical talker's voice with a high relative clause attachment (e.g., *The uncle of the girl who will ride the motorbike is from France*) and another critical talker's voice with a low attachment (e.g., *The uncle of the girl who will ride the carrousel is from France*). An example clipart scene showed a little girl, a man, a carrousel, and a motorbike, a glass of beer, and a jar of sweets. A neutral talker produced both attachments equally often during training. During testing, participants saw the same scene but heard sentences involving other depicted objects (e.g., *The uncle of the girl who will taste the sweets / beer*…). They began to inspect the correct target object (e.g., the sweets/beer) during the verb *ride* for the critical (e.g., low-/high-attachment) talker but not for the neutral talker. These looks occurred as quickly as for simple SVO sentences in which participants' long-term linguistic and world knowledge alone guided visual anticipation of the target object (e.g., *The girl will taste the sweets*; see Kamide et al. 2003b). It is possible that both the scenes and the linguistic context in Kamide (2012) provided a sufficiently rich event context that enabled the observed rapid effects of the short-term talker–sentence structure associations on structural choice.

Overall, the results reviewed in Section 2 suggest that not all comprehension processes manifest themselves immediately in the gaze record relative to when relevant information becomes available in the input: delays in the gaze record reflected difficulty associated with the computation of sentence structure, of scalar implicature, and – to some extent – recently learned world–language associations. The review in Section 3 highlights further that a range of lexical world–language relations (e.g., semantic associations of *piano* with a trumpet) can temporarily modulate our visual attention and language comprehension but that most attention still goes to referents (e.g., the piano). While this may be obvious in an example that contrasts a piano with a trumpet as participants listen *to piano*, the next section reviews evidence for the view that a referential priority also plays an important role at the sentence level, that it emerges in many different studies (and not just for noun–object but also for verb–action relations), and that it emerges even when pitted against other cues and strong probabilistic biases.

3. Priority of Referential over other World–Language Relations

In this section, we will discuss relevant evidence in favor of and against a referential priority from both adult and child language comprehension and discuss to which extent the priority can be accommodated by probabilistic biases alone.

3.1. ADULT LANGUAGE COMPREHENSION

A first study on how information type affects sentence comprehension in real time contrasted the relation of a verb ('spies-on') to its situation-immediate referent (a spying action performed by a wizard) with its relation to a stereotypical agent (a detective, Knoeferle and Crocker 2006). In the utterance *Den Piloten bespitzelt gleich…* ('The pilot (obj) spies-on soon'), the verb could either be related to a depicted spying action (and guide attention to its agent) or be related to a nearby stereotypical agent (a detective, depicted as performing an unrelated action). In this experimental situation, participants inspected the depicted action and its agent more often during the verb, thus prioritizing verb–action reference over expectations of what a stereotypical agent might do next (Knoeferle and Crocker 2006).

In another study, participants listened to German sentences (e.g., *Der Kellner poliert…* 'The waiter polish…') in which the verb stem was ambiguous between referencing a recently inspected action (polishing candelabra) or an equally plausible future polishing action involving another depicted object (e.g., crystal glasses, Knoeferle and Crocker 2007, Experiment 3). The sentence contained a tense manipulation (the last letter of the verb *poliert-e* and ensuing adverb either cued the recent event, or the verb was in the present tense and followed by a future tense adverb, resulting in a future tense sentence meaning). The sentence ended by mentioning either the recent (the candelabra) or the future (the crystal glasses) target object such that mention of these targets was balanced within the study. Participants' gaze pattern during 'polishes/d' and the ensuing adverb ('soon / recently') revealed a preference to relate the verb to the recently inspected action and its location (they inspected the location of the recent action and its co-located target object, the candelabra, more than the crystal glasses that could be polished next).

More recent studies have extended the latter finding to real-world actions and have provided evidence for the view that the recent-event preference cannot be accommodated by long-term frequency biases of the verbs or by the higher frequency of the recent actions. Knoeferle and colleagues (2011, Experiment 2, see Figure 3) examined whether within-experiment frequency biases (of a 'recent' versus 'future' real-world action) caused the preferred reliance on recent events observed by Knoeferle and Crocker (2007, Experiment 3). In the latter clipart studies, both recent and future actions were mentioned equally often, but people had always seen one 'recent' action per trial, prior to sentence comprehension, while the future event was never acted out. This within-experiment frequency bias towards recent events may have caused the preferred reliance on recent (vs. future) events. However, the preference persisted even when

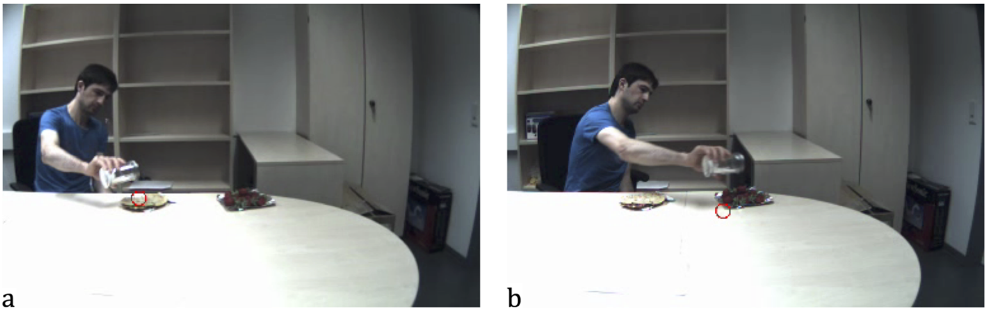

Fig 3. Snapshots from the real-world setting in Knoeferle et al. (2011) from the participant's perspective (the red circle represents the participant's gaze and was not visible to participants during the study). Participants first saw the experimenter sugar the pancakes (picture a). Then they heard either Der Versuchsleiter zuckerte kürzlich die Pfannkuchen 'The experimenter sugared recently the pancakes' or Der Versuchsleiter zuckert demnächst die Erdbeeren 'The experimenter sugars soon the strawberries'. After sentence presentation, they saw the experimenter sugar the strawberries (picture b, Knoeferle et al. 2011, Experiment 2).

participants saw a balanced frequency distribution of 'recent' and 'future' real-world actions (each trial showed one event before and the other after sentence presentation, referred to in the past and future tenses respectively). These findings suggest that the preferential inspection of the recent action target was not caused by an imbalance in the frequency of the performed events. Abashidze et al. (2014) used the same design as Knoeferle et al. (2011) but increased the frequency of the future event to 75% in one study and to 88% in another study. The frequency bias towards the future events resulted in an earlier effect of tense (and rise of looks to the future target event post-verbally) relative to previous studies in which the frequency of recent and future events was balanced (e.g., Experiment 2 in Knoeferle et al. 2011). However, starting at the verb and throughout the sentence, the same overall preference of more looks to the target of the recent than future event was replicated even with this strong bias towards future events.

In summary, world–language relations appear to differ in the extent to which they inform visual attention and language comprehension: all else being equal, when relating verbs to the visual context, action depictions (or their recent locations and co-located targets) are preferred over either a stereotypically plausible agent of the verb or a future event (and the latter holds even when future events are much more frequent).

Results from an experiment by Altmann and Kamide (2009) might appear to contradict the findings by Knoeferle and colleagues at first glance. In a visual world eye-tracking study, the authors examined whether listeners' visual attention could reflect their representation of described events even when those events were not depicted and when the depiction showed another state of the world. A woman was described as either moving a (depicted) glass to a (depicted) table or leaving it on the floor. Then she was described as pouring wine into the glass (…*pour the wine carefully into the glass*). The main question was whether people would direct their eye gaze to the table more often when they had been told that the glass had been moved there compared with when it had been described as remaining on the floor (the depicted glass never changed its location). Participants indeed inspected the table more often after hearing *pour* when the context had (vs. had not) previously described the glass as having been moved there.

Given this finding, it might be tempting to emphasize (as the authors have done) that visual attention is predominantly guided by a mental model of the described and imagined world. This is without doubt interesting, but one should also highlight that while the narrated world (glass on the table vs. on the floor) subtly modulated visual attention, participants made overall more eye fixations to the glass on the floor than to the table (even when the glass was described as having been moved to the table). This referential effect was only eliminated when the visual context was removed during comprehension (thus potentially decreasing its relevance for the comprehension situation, see also Knoeferle and Crocker 2007, Experiment 2, for relevant discussion on the role of working memory; see also Zwaan 2014).

A strong referential priority is also apparent in the results from Scheepers, Keller, and Lapata's (2008) study on 'coercion' phenomena, that is, how verbs such as *start* in *The artist started the painting* come to mean 'started to paint'. The verb referred to an action (e.g., *The artist painted / analyzed the flowery picture…*) or was abstract (e.g., *The artist started the flowery picture…*), and if concrete, it was either preferred (*painted*) or dispreferred (*analyzed*) in the sentence context. Eye fixations on depicted instruments (a paintbrush vs. a magnifying glass) suggested that visual attention to target objects, and thus language processing, was faster for concrete referential than metonymic verbs such as 'start'.

The view that reference is prioritized over other lexical associations receives further support from a study on the processing of concrete relative to abstract words by Duñabeitia, Avilés, Afonso, Scheepers, and Carreiras (2009). When world–language relations were associative in nature (abstract: Spanish 'smell' was associated with the picture of a nose; concrete: Spanish 'crib' was associated with the picture of a baby), abstract words elicited more and earlier looks to the associated target picture than concrete ones (from 200 to 400 ms after word onset), suggesting differences in the mental representations of associations for abstract relative to concrete words. Crucially, when the same pictures were named ('nose' and 'baby'), no such gaze differences emerged, and participants inspected these referents more than they had inspected them when hearing associated nouns such as 'smell' or 'crib'. Thus, referential relations ('nose' – picture of a nose) elicited more inspection of the target than did associative relations (e.g., 'smell' – picture of a nose).

It is worth noting that words guide visual attention only to communicatively relevant aspects of the visual context. In Chambers and San Juan (2008, Experiment 2), participants in one 'unique' condition were instructed to move an object (e.g., a truck dubbed 'guitars', *Send the guitars to the store…*), and then they were asked to return it to its original location, area 7 (*Now return the guitars to area 7*). In a second 'non-unique' condition, the instructions involved two different trucks prior to the critical sentence (*Now return the guitars…*). In a third 'incidental' condition, participants had to move the guitars to another area, but to do so, they had to move another truck (the sweets) out of the way. Thus, only one of these two moves was relevant to the conversation. In this article, participants inspected the guitars as often in the unique as in the incidental condition and more in both of these than in the non-unique condition. The authors interpreted this finding as evidence that objects matching the verb semantically (the incidentally moved sweets truck) could be ignored if their move had not been explicitly instructed. While this at first glance seems to provide evidence against a referential preference, there are at least two points to consider. First, looks to the incidentally moved sweets truck were not reported. If the sweets truck also elicited substantial inspection, this would support the view that any recent action increases attention to its target. Moreover, the recent-event preference is not a 'dumb' mechanism, that is, the account clarifies that words increase visual attention to *relevant* aspects of the scene. An incidentally moved object would appear to be implicitly excluded since it is not directly relevant to processing the instruction (see Knoeferle and Crocker, p. 542).

3.2. CHILD LANGUAGE COMPREHENSION

We can also assess the importance of referential world–language relations by examining children's language learning and comprehension. If we assume continuity of language comprehension from early childhood into adulthood, then a central role of visual context in adult comprehension would predict clear referential effects at developmental stages (see Knoeferle 2015b). It's not implausible that children have a preference for establishing reference since even young children rapidly fixate the picture of a word they have recognized (see Fernald, McRoberts, and Swingley 2001; Hollich, Hirsh-Pasek, and Golinkoff 2000; Swingley, Pinto, and Fernald 1999).

The role of the visual context for child language comprehension has, however, been questioned. Trueswell, Sekerina, Hill, and Logrip (1999) examined visual context effects in young adults relative to 5-year-olds. In the sentence fragment *Put the frog on the napkin…*, the prepositional phrase *on the napkin* could either modify *the frog*, indicating its location in the visual context, or the verb phrase, specifying the destination of the action. When there was only one frog, the adults preferred the destination interpretation (inspecting an empty napkin to which the frog could be moved more than the frog on the napkin as they heard *napkin*). By contrast, when there were two frogs and only one of them was on a napkin, *on the napkin* could serve to identify the correct frog (see also Spivey Tanenhaus, Eberhard, and Sedivy 2002; Tanenhaus et al. 1995). The adults thus rapidly adopted the location interpretation and inspected the frog located on a napkin and hardly inspected the empty napkin. However, when 5-year-olds in the study by Trueswell et al. listened to this sentence, they frequently inspected the potential destination of the action (the empty napkin) instead of the frog on the napkin, even when the context supported the location interpretation (one of two frogs was on a napkin).

At first glance, this may look as if the children, unlike the adults, failed to use the referential visual context. However, when we more closely consider the processes underlying children's gaze behavior, a referential strategy emerges. If children pursued a referential strategy when hearing *on the napkin*, they would look at the single empty napkin (vs. another object on a napkin). Such a strategy would precisely direct their gaze to the incorrect destination (the empty napkin; see Zhang and Knoeferle 2012 for discussion).

Indeed, when relevant visual information was directly referenced, visual context (depicted agent–action–patient events) rapidly affected children's real-time visual attention and incremental thematic role assignment. In a study by Zhang and Knoeferle (2012), children in one condition inspected three characters (e.g., a bull, a bear, and a worm), and in addition, the scene depicted action events between them (e.g., a bull pushing a bear). In another condition, only the three characters were depicted. As they inspected one of these scenes, the children listened to a German SVO or OVS sentence (e.g., *Den Bär schubst sogleich der Stier*, 'The bear (obj.) pushes immediately the bull (subj)'). Post-sentence, when any event depiction had been removed, they responded to a question about thematic role relations (e.g., *Who pushed*?). Seeing event depictions during comprehension increased children's post-sentence accuracy on the difficult OVS sentences by around 18% (without events their accuracy was at chance despite unambiguous case marking; see Dittmar, Abbot-Smith, Lieven, and Tomasello 2008). Like the adults, children initially inspected the bear during (*Den Bär*, 'the bear (obj)'), and once the verb had identified an event in which the bear was the patient ('pushes'), they anticipated the agent of the pushing event (the bull). This pattern further varied depending on children's accuracy and working memory scores such that the time course of high (but not low) working memory children's eye fixations resembled that of adults. Children can thus rapidly integrate verb-mediated depicted actions and their associated agents for performing incremental thematic role assignment and syntactic structuring. Whether this result extends to the disambiguation of local structural ambiguity remains to be seen.

Overall thus, referential relations affect object-directed visual attention more (and sometimes even earlier; see Scheepers et al. 2008) than other world–language relations during sentence processing. Existing evidence further supports the view that this referential priority can be attenuated by our short-term experience (e.g., Abashidze et al. 2014) or by other situation-immediate cues (e.g., an actor's eye gaze, Abashidze et al. 2015). However, even when we pitch these other cues (e.g., an actor's gaze or stereotypical thematic role knowledge) or factors (short-term frequency of a cue) against the referential preference, the latter replicates. This preference moreover shines through in studies examining a diverse range of topics – from the investigation of narrated events and their mental models to the organization of semantic memory and coercion phenomena.

4. Summary and Conclusions

The reviewed results highlight the incremental modulation of visual attention through phonological, semantic, syntactic, and pragmatic processes. Visual attention to objects was delayed for non-canonical relative to canonical structures, for processes such as computing scalar implicature (vs. semantic meaning) and, to some extent, when accessing learned, short-term associations (vs. long-term linguistic and world knowledge). These findings suggest that the close time-locking of gaze to the utterance interpretation varies with the implicated mental processes (delays can, for instance, reflect difficulty with non-canonical structure, time needed to complete complex inferential processes, or time needed to retrieve short-term associations). However, a sufficiently rich context was able to eliminate this sort of delay at least for short-term associations between a talker and syntactic structures.

While this sort of context variation highlights the flexibility of the system, should we assume that visual attention varies unsystematically with context? This does not seem to be the case. All else being equal, one systematic aspect in the deployment of visual attention is the existence of a referential priority: when a noun referred to an object, comprehenders were more likely to inspect that object than other, lexico-semantically associated objects. When a verb referred to an action, people were more likely to inspect the action and its agent than another, stereotypically plausible agent. Concerning the linking hypotheses between fixation probabilities and language–world relations, referential relations thus seem to take priority over other world–language relations. Existing accounts have accordingly postulated that comprehenders first check the scene for referential relations rather than solely relying on linguistic/world knowledge (Knoeferle and Crocker 2006).

Current research in our laboratory is pursuing this topic further by comparing referential against other world–language relations. For instance, we compared a verb-mediated depicted action with speaker gaze effects in a crossed, two-by-two design (actions cueing a target were present vs. absent; the speaker gazed at the target object or was obscured; Kreysa, Knoeferle, and Nunnemann 2014). By directly comparing the effects of referential cues and other world–language relations, we can begin to characterize their relative contribution to situated language comprehension – perhaps in the form of a hierarchy.

The present article has emphasized a psycholinguistic analysis of how different comprehension processes and world–language relations affect visual attention. Another level of analysis would be in terms of probabilistic information. In this regard, we note that long-term linguistic experience did not accommodate the observed referential priority at least for verb–action relations, and the priority of recent events was not eliminated by short-term frequency manipulations (i.e., of a higher frequency for future than past sentences and events). However, existing studies have not yet controlled the long-term frequency of different world–language relations (i.e., long-term, referential world–language relations may be more frequent than other world–language relations). For instance, the observed referential priority could be accommodated probabilistically if an action verb referred more often to an action in the immediate environment than to an action that a stereotypical agent might perform next. For nouns, a probabilistic account would predict that they occur more often with referents than with other, associated objects.

From a theoretical viewpoint, we will want to complement a probabilistic level of analysis with an explanatory psycholinguistic account (detailing why comprehenders prioritize referential relations). For instance, we might reason that this behavior in young adults has a developmental basis and that the importance of referential relations for word learning in infancy is at the origin of the referential priority in young adults' language comprehension (Knoeferle 2015b). An alternative (or complementary) explanation is epistemic in nature. Comprehenders might prefer to relate a referential expression (e.g., a verb) to its referent because that relation can be immediately verified. By contrast, absent referents such as future actions cannot be verified, resulting in greater uncertainty for the unfolding interpretation (MacFarlane 2003; Staub and Clifton 2011; see Abashidze et al. 2014).

While we have emphasized the importance of referential relations in the present article, other cognitive factors (e.g., a comprehender's cognitive resources; see Knoeferle and Crocker 2007), the communicative relevance of cues (e.g., Chambers and San Juan 2008), and the timing and complexity of stimuli (e.g., Ferreira, Foucart, and Engelhardt 2013) have also been shown to modulate the interrogation of the visual context and visual context effects (see also Knoeferle, Urbach, and Kutas 2011, 2014). Future research could profitably examine to which extent the relative contribution of different world–language relations to comprehension varies as a function of these factors. Overall, we argue that investigations such as the ones reviewed in Section 3 can offer important constraints – in the form of a systematic characterization of the relative contribution of distinct cues – for a processing account of visually situated language comprehension.

Acknowledgement

This research was supported by the Bielefeld University Excellence Center for Cognitive Interaction Technology (277, German Research Council, DFG). We thank the reviewers for their helpful comments. This research was funded by the Excellence Center 277 'Cognitive Interaction Technology' (German Research Council, DFG), and by the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement n°316748 (LanPercept).

Short Biographies

Pia Knoeferle holds two undergraduate degrees (in English Philology, Romance Studies, and Philosophy, as well as in English Studies, French, and Physical Education) from Regensburg University, Germany. She received a Philosophical Doctorate in 2005 ("summa cum laude") as well as the Eduard-Martin Prize for outstanding dissertations from Saarland University, Germany. While on a postdoctoral fellowship at UC San Diego (2007–2008, funded by the German Research Council), she founded a workshop on "Embodied and situated language processing 2007" which has since taken place in Rotterdam (2009), San Diego (2010), Bielefeld (ZiF, 2011), Northumbria University (2012), Potsdam University (2013), Rotterdam (2014), and Lyon (2015). She was first Assistant and then Associate Professor for "Language and Cognition" at Bielefeld University (Department of Linguistics and Cognitive Interaction Technology Excellence Center, 2009–2015) and is now Full Professor of "German Linguistics: Psycholinguistics" at the Humboldt University in Berlin, Germany. She is on the editorial board of Frontiers in Language Sciences and Frontiers in Cognition, and on the board of Reviewers for Cognitive Science. Dr. Knoeferle examines the real-time interaction of information from non-linguistic context, visual attention, and incremental language comprehension. In her research she has relied upon eye tracking and event-related brain potentials, as well as, collaboratively, computational modeling. Some of the current topics in her group are how a speaker's gaze directs a listener visual attention; how comprehenders use emotional facial expressions during language comprehension; how spatial expressions can guide visual attention; and effects of depicted events on visual attention and spoken language comprehension in children and adults.

Ernesto Guerra holds a Joint Master's degree from Potsdam University (Germany) and Groningen University (Netherlands), and a Doctoral degree from Bielefeld University, Germany. Following his doctorate, he was a postdoctoral scholar at the Max Planck Institute for Psycholinguistics, Nijmegen before taking up his current position as a Postdoctoral Fellow at the Experimental Psycholinguistics Lab & Interdisciplinary Center for Intercultural and Indigenous Studies, Pontificia Universidad Católica de Chile. His expertise lies in the area of eye tracking during language comprehension and his research interests include the role of perceptual cues in sentence processing and individual differences in text comprehension.

Notes

* Correspondence address: Pia Knoeferle, Department of German Language and Linguistics, Humboldt University, Berlin, Unter den Linden 6, 10099 Berlin. E-mail: pia.knoeferle@hu-berlin.de

[1] Note that Cooper (1974) had already discovered the connection between spoken comprehension and visual attention, but his findings did not impact mainstream psychological and cognitive science research at the time.

[2] A linking hypothesis is an assumption about how data patterns relate to cognitive processes.

[3] Note that the reviewed studies differed in task. In Allopenna et al. (1998), for instance, participants were instructed to click on and move one out of four displayed objects to another location on a virtual grid, and the same was true for the experiments by Creel et al. (2008). In Huettig and Altmann (2005), by contrast, participants did not perform an explicit task but listened passively to utterances as they inspected an array of four objects. While some task differences may substantially modulate gaze pattern and language comprehension processes (e.g., Knoeferle and Kreysa 2012; Salverda, Brown, and Tanenhaus 2011), not all of them will (e.g., Altmann and Kamide 1999; Burigo and Knoeferle 2015).

[4] A positive deviation in mean amplitude ERPs approximately 600 ms after the onset of a stimulus has been observed following syntactic violations and in the case of structural revision. This brain response is dubbed a P600 (also syntactic positive shift, e.g., Hagoort, Brown, and Groothusen 1993; Osterhout and Holcomb 1992, 1993).

Works Cited

Abashidze, Dato, Pia Knoeferle, and Maria Nella Carminati. 2014. How robust is the recent event preference? *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, ed. by Paul Bello, Marcello Guarini, Marjorie McShane and Brian Scassellati, 92–7. Austin, TX: Cognitive Science Society.

——. 2015. Eye-tracking situated language comprehension: immediate actor gaze versus recent action events. Proceedings of the 37th Annual Meeting of the Cognitive Science Society, ed. by David C. Noelle, Rick Dale, Anne S. Warlaumont, Jeff Yoshimi, Teenie Matlock, Carolyn D. Jennings and Paul P. Maglio, 31–6. Austin, TX: Cognitive Science Society.

Allopenna, Paul. D., James S. Magnuson, and Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language* 38. 419–39.

Altmann, Gerry T. 2011. Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica* 137. 190–200.

Altmann, Gerry T., and Yuki Kamide. 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73. 247–64.

——. 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.

——. 2009. Discourse-mediation of the mapping between language and the visual world: eye movements and mental representation. *Cognition* 111. 55–71.

Altmann, Gerry T, and Jelena Mirković. 2009. Incrementality and prediction in human sentence processing. *Cognitive Science* 33. 583–609.

Barrett, Sarah E., and Michael D. Rugg. 1990. Event-related potentials and the semantic matching of pictures. *Brain and Cognition* 14. 201–12.

Ben-David, Boaz M., Craig G. Chambers, Meredyth Daneman, M. Kathleen Pichora-Fuller, Eyal M. Reingold, and Bruce A. Schneider. 2011. Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *Journal of Speech, Language, and Hearing Research* 54. 243–62.

Burigo, Michele, and Pia Knoeferle. 2015. Visual attention during spatial language comprehension. *PLoS ONE* 10. 1–21, DOI:10.1371/journal.pone.0115758.

Chambers, Craig G., and Valerie San Juan. 2008. Perception and presupposition in real-time language comprehension: insights from anticipatory processing. *Cognition* 108. 26–50.

Clark, Herbert A., and William G. Chase. 1972. On the process of comparing sentences against pictures. *Cognitive Psychology* 3. 472–517.

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.

Creel, Sarah C., Richard N. Aslin, and Michael K. Tanenhaus. 2008. Heeding the voice of experience: the role of talker variation in lexical access. *Cognition* 106. 633–64.

Dahan, Delphine. 2010. The time course of interpretation in speech comprehension. *Current Directions in Psychological Science* 19. 121–26.

Dahan, Delphine, and Michael K. Tanenhaus. 2005. Looking at the rope when looking for the snake: conceptually mediated eye-movements during spoken-word recognition. *Psychonomic Bulletin & Review* 12. 453–59.

Deno, Stanley L. 1968. Effects of words and pictures as stimuli in learning language equivalents. *Journal of Educational Psychology* 59. 202–6.

Dittmar, Miriam, Kirsten Abbot-Smith, Elena Lieven, and Michael Tomasello. 2008. German children's comprehension of word order and case marking in causative sentences. *Child Development* 79. 1152–67.

Duñabeitia, Jon A., Alberto Avilés, Olivia Afonso, Christoph Scheepers, and Manuel Carreiras. 2009. Qualitative differences in the representation of abstract versus concrete words: evidence from the visual-world paradigm. *Cognition* 110. 284–92.

Fernald, A., G. W. Roberts, and D. Swingley. 2001. Infants' developing competence in recognizing and understanding words in fluent speech. *Approaches to bootstrapping: phonological, lexical, syntactic, and neurophysiological aspects of early language acquisition*. Volume I, ed. by Jürgen Weissenborn and Barbara Höhle, 97–123. Amsterdam, The Netherlands: John Benjamins.

Ferreira, Fernanda, Alice Foucart, and Paul E. Engelhardt. 2013. Language processing in the visual world: effects of preview, visual complexity, and prediction. *Journal of Memory and Language* 69. 165–182.

Ganis, Giorgio, Marta Kutas, and Martin I. Sereno. 1996. The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience* 8. 89–106.

Gough, Philip B. 1966. The verification of sentences: the effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior* 5. 492–6.

Grodner, D., Klein, N., Carbary, K., and Tanenhaus, M. (2010). 'Some', and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment. *Cognition* 116. 42–55.

Hagoort, Peter, Colin Brown, and Jolanda Groothusen. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes* 8. 439–83.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics. 159–66. Pittsburg, PA.

Hemforth, Barbara. 1993. *Kognitives Parsing: Repräsentation und Verarbeitung grammatischen Wissens*. Sankt Augustin: Infix.

Hollich, George J., Kathy Hirsh-Pasek, and Roberta M. Golinkoff. 2000. Breaking the language barrier: an emergentist coalition model of the origins of word learning. *Monographs of the Society for Research in Child Development* 65. 1–16.

Huang, Yi Ting, and Jesse Snedeker. 2009. Semantic meaning and pragmatic interpretation in 5-Year-Olds: evidence from real-time spoken language comprehension. *Developmental Psychology* 45. 1723–39.

Huettig, Falk and Gerry T. M. Altmann. 2005. Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96. B23–32.

Huettig, Falk, Joost Rommers, and Antje S. Meyer. 2011. Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica* 137. 151–71.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Just, Marcel A., and Patricia A. Carpenter. 1971. Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior* 10. 244–53.

Kaiser, Elsi and John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94. 113–47.

Kamide, Yuki. 2008. Anticipatory processes in sentence processing. *Language and Linguistics Compass* 2. 647–670.

——. 2012. Learning individual talkers' structural preferences. *Cognition* 124. 66–71.

Kamide, Yuki, Gerry T. M. Altmann, and Sarah L. Haywood. 2003b. The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *Journal of Memory and Language* 49. 133–56.

Kamide, Yuki, Christoph Scheepers, and Gerry T. M. Altmann. 2003a. Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research* 32. 37–55.

Kellogg, Gillian S., and Michael J. A. Howe. 1971. Using words and pictures in foreign language learning. *Alberta Journal of Educational Research* 17. 89–94.

Knoeferle, Pia. 2015a. Language comprehension in rich non-linguistic contexts: combining eye-tracking and event-related brain potentials. *Cognitive neuroscience of natural language use*, ed. by Roel M. Willems, 77–100. Cambridge, UK: Cambridge University Press.

——. 2015b Visually situated language comprehension in children and in adults. In Attention and vision in language processing, ed. by Ramesh K. Mishra, Narayanan Srinivasan and Falk Huettig, 57–77. Berlin: Springer.

Knoeferle, Pia, and Maria Nella Carminati, Dato Abashidze, and Kai Essig. 2011. Preferential inspection of recent real-world events over future events: evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology* 2. 1–12.

Knoeferle, Pia, and Matthew W. Crocker. 2006. The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science* 30. 481–529.

——. 2007. The influence of recent scene events on spoken comprehension: evidence from eye movements. *Journal of Memory and Language* 57. 519–43.

——. 2009. Constituent order and semantic parallelism in online comprehension: eye-tracking evidence from German. *The Quarterly Journal of Experimental Psychology* 62. 2338–71.

Knoeferle, Pia, Matthew W. Crocker, Christoph Scheepers and Martin J. Pickering. 2005. The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition* 95. 95–127.

Knoeferle, Pia, and Helene Kreysa. 2012. Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology* 3. 1–15.

Knoeferle, Pia, Thomas Urbach, Marta Kutas. 2011. Comprehending how visual context influences incremental sentence processing: insights from ERPs and picture-sentence verification. *Psychophysiology* 48. 495–506.

——. 2014. Different mechanisms for role relations versus verb–action congruence effects: evidence from ERPs in picture–sentence verification. *Acta Psychologica* 152. 133–48.

Kreysa, Helene, Pia Knoeferle, and Eva M. Nunnemann. 2014. Effects of speaker gaze versus depicted actions on visual attention during sentence comprehension. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, ed. by Paul Bello, Marcello Guarini, Marjorie McShane and Brian Scassellati, 2513–18. Austin, TX: Cognitive Science Society.

Kroll, Judith F., and Ann Corrigan. 1981. Strategies in sentence–picture verification: effects of an unexpected picture. *Journal of Verbal Learning and Verbal Behavior* 20. 515–31.

Kutas, Marta, and Kara D. Federmeier. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential ERP. *Annual Review of Psychology* 62. 621–46.

Kutas, Marta, and Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307. 161–3.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–77.

MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101. 676–703.

MacFarlane, John. 2003. Future contingents and relative truth. *The Philosophical Quarterly* 53. 321–336.

Marquer, Josette, and Maria Pereira. 1990. Reaction times in the study of strategies in sentence–picture verification: a reconsideration. *The Quarterly Journal of Experimental Psychology* 42. 147–68.

Mathews, Nancy N., Earl B. Hunt, and Colin M. MacLeod,. 1980. Strategy choice and strategy training in sentence–picture verification. *Journal of Verbal Learning and Verbal Behavior* 19. 515–48.

Matzke, Mike, Heinke Mai, Wido Nager, Jascha Rüsseler, and Thomas C. Münte. 2002. The costs of freedom: an ERP-study of non-canonical sentences. *Clinical Neuropsychology* 113. 844–52.

Moeser, Shannon, and Albert S. Bregman. 1972. Imagery and language acquisition. *Journal of Verbal Learning and Verbal Behavior* 12. 91–8.

Moeser, Shannon, and Joyce A. Olson. 1974. The role of reference in children's acquisition of a miniature artificial language. *Journal of Experimental Child Psychology* 17. 204–18.

Nigam, Arti, James E. Hoffman, and Robert F. Simons. 1992. N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience* 4. 15–22.

Osterhout, Lee, and Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31. 785–806.

——. 1993. Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes* 8. 413–37.

Paivio, Allan. 1971. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

——. 1986. Mental representations: a dual coding approach. Oxford, UK: Oxford University Press.

Pezdek, Kathy. 1977. Cross-modality semantic integration of sentence and picture memory. *Journal of Experimental Psychology: Human Learning and Memory* 3. 515–24.

Potter, Mary C., and Barbara A. Faulconer. 1975. Time to understand pictures and words. *Nature* 253. 437–38.

Potter, Mary C., C. Judith F. Kroll, Betsy Yachzel, Elizabeth Carpenter, and Janet Sherman. 1986. Pictures in sentences: understanding without words. *Journal of Experimental Psychology: General* 115. 281–94.

Salverda, Anne Pier, M. Brown, and Michael K. Tanenhaus. 2011. A goal-based perspective on eye movements in visual world studies. *Acta Psychologica* 137. 172–180.

Salverda, Anne Pier, Delphine Dahan, James M. McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90. 51–89.

Scheepers, Christoph, Frank Keller, and Mirella Lapata. 2008. Evidence for serial coercion: a time course analysis using the visual-world paradigm. *Cognitive Psychology* 56. 1–29.

Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers, and Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71. 109–47.

Shepard, Roger N. 1967. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior* 6. 153–63.

Simons, Daniel J. 1996. In sight, out of mind: when objects representation fail. *Psychological Science* 7. 301–305.

Spivey, Michael J., Michael K. Tanenhaus, Kathleen M. Eberhard, and Julie C. Sedivy. 2002. Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology* 45. 447–81.

Staub, A., and Clifton, C., Jr. 2011. Processing effects of an indeterminate future: evidence from self-paced reading. *University of Massachusetts Occasional Papers in Linguistics*, Vol. 38, ed. by Jesse A. Harris and Margaret Grant, 131–140. Amherst, MA: GLSA Publishing.

Swingley, D., John P. Pinto, and A. Fernald. 1999. Continuous processing in word recognition at 24 months. *Cognition* 71. 73–108.

Tanenhaus, Michael K., John M. Carroll, and Thomas G. Bever. 1976. Sentence–picture verification models as theories of sentence comprehension: a critique of carpenter and just. *Psychological Review* 83. 310–17.

Tanenhaus, Michael K., James S. Magnuson, Delphine Dahan, and Craig Chambers. 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, Michael K., Michael J. Spivey-Knowton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–4.

Tanenhaus, Michael K., and John C. Trueswell. 2006. Eye movements and spoken language comprehension. *Handbook of Psycholinguistics*, ed. by Matthew Traxler and Morton A. Gernsbacher, 863–900. New York, NY: Academic Press.

Tomasello, Michael, and Michael J. Farrar. 1986. Joint attention and early language. *Child Development* 57. 1454–63.

Trueswell, John C., Irina Sekerina, Nicole M. Hill, Marian L. Logrip. 1999. The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition* 73. 89–134.

Trueswell, John C., and Michael K. Tanenhaus. 1994. Toward a lexical framework of constraint-based syntactic ambiguity resolution. *Perspectives on sentence processing*, ed. by Lyn Frazier and Keith Rayner, 155–179. Lawrence Erlbaum Associates.

Weber, Andrea, Martine Grice, and Matthew W. Crocker. 2006. The role of prosody in the interpretation of structural ambiguities: a study of anticipatory eye movements. *Cognition* 99. B63–B72.

Whitehurst, Grover J., David S. Arnold, Jeffery N. Epstein, Andrea L. Angell, Meagan Smith, Janet E. Fischel. 1994. A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology* 30. 679–89.

Whitehurst, Grover J., Francine Falco, Christopher J. Lonigan, Janet E. Fischel, Barbara D. DeBaryshe, Marta Valdez-Menchaca, Marie Caulfield. 1988. Accelerating language development through picture book reading. *Developmental Psychology* 24, 552–59.

Yee, Eiling, Sheila E. Blumstein, and Julie C. Sedivy. 2008. Lexical–semantic activation in Broca's and Wernicke's aphasia: evidence from eye movements. *Journal of Cognitive Neuroscience* 20. 592–612.

Yee, Eiling, and Julie C. Sedivy. 2006. Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32. 1–14.

Zhang, Lu, and Pia Knoeferle. 2012. Visual context effects on thematic role assignment in children versus adults: evidence from eye tracking in German. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, ed. by Naomi Miyake, David Peebles, and Richard P. Cooper, 2593–8. Austin, TX: Cognitive Science Society.

Zwaan, Rolf A. 2014. Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences* 18. 229–34.