Knoeferle, P. (to appear). Language comprehension in rich non-linguistic contexts: combining eye tracking and event-related brain potentials. In: Roel, Williams (Ed.). Towards a cognitive neuroscience of natural language use. Cambridge: Cambridge University Press.

Language comprehension in rich non-linguistic contexts: combining eye tracking and event-related brain potentials

Pia Knoeferle (knoeferl@cit-ec.uni-bielefeld.de)
Cognitive Interaction Technology Excellence Center (CITEC)
Bielefeld University, Germany

Abstract

The present chapter reviews the literature on visually situated language comprehension against the background that most theories of real-time sentence comprehension have ignored rich non-linguistic contexts. However, listeners' eye movements to objects during spoken language comprehension, as well as their event-related brain potentials (ERPs) have revealed that non-linguistic cues play an important role for real-time comprehension. In fact, referential processes are rapid and central in visually situated spoken language comprehension and even abstract words are rapidly grounded in objects through semantic associations. Similar ERP responses for non-linguistic and linguistic effects on comprehension suggest these two information sources are on a par in informing language comprehension. ERPs further revealed that non-linguistic cues affect lexical-semantic as well as compositional processes, thus further cementing the role of rich non-linguistic context in language comprehension. However, there is also considerable ambiguity in the linking between comprehension processes and each of these two measures (eye movements and ERPs). Combining eye-tracking and event-related brain potentials would improve the interpretation of individual measures and thus insights into visually-situated language comprehension.

Knoeferle, P. (to appear). Language comprehension in rich non-linguistic contexts: combining eye tracking and event-related brain potentials. In: Roel, Williams (Ed.). Towards a cognitive neuroscience of natural language use. Cambridge: Cambridge University Press.

1. Introduction

Much of our everyday language use occurs in contextually rich settings. This is true, for instance, when we select a tram ticket to go to work and follow the instructions of a vending machine; when we read the paper; or when we buy a croissant at the corner bakery. At the vending machine, for instance, we can use verbal labels such as "day-ticket" together with depictions of zones on the city map to understand which kind of ticket we are buying and where it is valid. In the bakery, we can gesture and point to a pastry if we don't know its name, and if we see the baker select a pastry that we don't like, we can ask him to give us another one. In fact, if the baker sees us scowl when he selects one of the smaller pastries, he may well pause, re-consider and hand us a larger one. Overall thus, perceived actions, object-based gaze, gestures, and facial expressions constitute a rich context for, and contribute relevant information to, our everyday communication.

1.1 Language-centricity in theories of sentence comprehension

While a view of comprehension as situated in a rich context seems intuitively plausible and appealing, this is not what has shaped psycholinguistic theorizing on real-time language comprehension. From the 1970s and well into the 1990s, a "language-centric" view has dominated theory formation and empirical research (e.g., Altmann & Steedman, 1988; Crocker, 1996; Forster, 1979; Frazier & Fodor, 1979; Frazier & Clifton, 1996; Gorrell, 1995; MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Trueswell & Tanenhaus, 1994). Without going into too much detail, early accounts of sentence comprehension were syntax-centric and accommodated structural decisions at choice points through principles such as syntactic simplicity and the use of purely structural rules (e.g., Frazier & Fodor, 1979). However, these accounts struggled to accommodate the rapid effects of lexical-semantic information on syntactic structure building. Accordingly theorizing turned to the lexicon as an important source of grammatical knowledge (e.g., Macdonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994) and to probabilistic information (e.g., Crocker & Brants, 2000; Mitchell et al., 1995; Spivey-Knowlton, & Tanenhaus, 1998). Recent approaches have accommodated comprehension by appealing to the likelihood of words in context (e.g., Hale, 2003; Levy 2008). While the above language-centric accounts have been shaped by reading times or eye-tracking data, others have been shaped by ERP results alone. Friederici (2002), for instance, proposed a neurocognitive model comprising three serial stages in sentence processing. A related, argument-dependency model assumes three hierarchically ordered stages that allow for some parallelism (Bornkessel & Schlesewsky, 2006). By comparison with the latter two models, a neuro-cognitive model based on unification grammar assumes parallel competition of syntactic and lexical structures without a separate initial phrase structure stage (Hagoort, 2003).

In sum, the above accounts and frameworks have all adopted a language-centric approach to comprehension. In line with this, they contribute valuable insights into a range of semantic and syntactic processes. Crucially however, none of them makes any predictions about how comprehension proceeds when language users can attend to and recruit all sorts of information from the immediate non-linguistic environment.

The present chapter takes the view that the latter is precisely among the situations in which we should be able to accommodate how comprehension proceeds, and how it benefits from non-linguistic relationships. Ultimately we want to model comprehension in all kinds of situations, for instance when only language is relevant, when text and pictures matter, and when speech relates (more or less) to objects and dynamically unfolding events (of which more below). This chapter reviews how the combination of continuous measures with rich non-linguistic contexts has paved the way towards examining 'visually-situated' language processing (i.e., language processing in situations when non-linguistic visual cues are relevant for comprehension). It further summarizes key insights into visually situated language comprehension from both eye-tracking and ERP studies. In the process, the chapter discusses issues concerning the linking hypotheses[1] in visually situated language comprehension and argues for combining these two measures to improve the interpretation of each individual measure.

1.2. Visually situated language comprehension: methodological advances and tasks

The observation that language-processing models are language-centric pertains specifically to real-time comprehension. For investigating the timing of events in language comprehension, researchers have relied upon continuous recordings of

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

either comprehenders' eye movements or their ERPs during text reading. The focus on reading and on linguistic theory arguably entailed a focus on linguistic contexts (sentences). By contrast, early research in cognitive psychology has examined language in richer contexts. One strand of research has examined picture-sentence verification (e.g., Carpenter & Just, 1976; Clark & Chase, 1972; Gough, 1965), another the nature of the mental representations underlying the processing of pictorial and linguistic stimuli (e.g. Potter et al., 1986; Potter & Kroll, 1987). Further approaches have modeled comprehension as perceived or imagined events (Johnson-Laird, 1981), or examined it in a dialogue context (e.g., Garrod & Anderson, 1987).

However, among these approaches, few have influenced theorizing in the area of real-time language processing (see Pickering & Garrod, 2004 for an exception), and most have had virtually no impact on psycholinguistic accounts of incremental language comprehension. This is arguably because the early cognitive-psychology research has largely relied upon non-continuous measures. Approaches such as picture-sentence verification were indeed criticized for not reflecting the mental processes implicated in real-time language comprehension (Tanenhaus, Carroll, & Bever, 1976). This criticism was motivated by the concern that picture-sentence verification - as indexed by post-sentence response latencies - could not reveal anything about moment-by-moment language comprehension.

What seems to have been overlooked, however, is that the criticism of picture-sentence verification mostly pertained to specific measures (e.g., speeded or post-sentence verification response times) but not to the task (e.g., verification) or research issue (picture-sentence verification, see Knoeferle, Urbach, & Kutas, 2011). In fact, when we employ continuous methods such as eye tracking and ERPs to study language processing in non-linguistic visual contexts, then a range of tasks appear suitable for providing insight into the time course and nature of language processing and into the interaction of comprehension processes with information from the non-linguistic context. Among these are tasks in which participants act out instructions on objects (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), listen for comprehension (e.g., Altmann & Kamide, 1999), judge sentence veracity (e.g., Guerra & Knoeferle, 2013), and verify picture-sentence congruence (e.g., Altmann & Kamide, 1999; Carminati & Knoeferle, 2013; Knoeferle et al., 2011; Vissers, Kolk, Van de Meerendonk, & Chwilla, 2008; Wassenaar & Hagoort, 2007).

These methodological advances (combining visually situated language tasks with continuous measures) tie in well with the goal of modeling comprehension in rich contexts. They have paved the way for addressing long-standing psycholinguistic questions as to whether linguistic and visual processes are or aren't informationally encapsulated (Tanenhaus et al., 1995, see Fodor, 1983), and as to which extent pictures and words are processed similarly (e.g., Ganis, Kutas, & Sereno, 1996). Psycholinguists have also examined how visual context affects syntactic structuring relative to lexical biases, with the conclusion that constraint-based interactive accounts of language processing can accommodate both visual-context and lexical effects (Novick, Thompson-Schill, & Trueswell, 2008). However, these accounts, just like the above language-centric accounts, do not aim to model the *nature* of the interplay between comprehension processes, (visual attention) and information from the non-linguistic visual context (see Altmann & Mirković, 2009; Crocker, Knoeferle, & Mayberry, 2010; Knoeferle & Crocker, 2006, 2007; Mayberry, Crocker, & Knoeferle, 2009).

Below I will first provide an overview of insights into visually situated language comprehension from eye-tracking data that addresses precisely this issue (2.1) and discuss the ambiguity in relating eye movements to comprehension processes (2.2). Section 2.3 introduces the possibility of complementing eye-tracking studies with ERP studies to remove ambiguity in the linking hypotheses and also discusses ERP results on the integration of co-speech gestures and language. The section addresses the interesting question of whether visual context affects comprehension in much the same way as our linguistic knowledge (2.3.1) and subsequently discusses ambiguity in the linking assumptions of ERPs (2.3.2). A final section (3) argues that the combination of eye tracking and ERPs can improve the interpretation of each individual measure in visually-situated language comprehension.

2. The interaction of language comprehension with non-linguistic cues

2.1 The time course and distribution of visual attention as cues to referential and semantic interpretation

Crucially, the monitoring of eye movements or ERPs during the presentation of visually situated sentences has laid the foundation for investigating how moment-by-moment language comprehension interacts with processing of (as well as visual attention in) the non-linguistic context. Results from eye-tracking studies suggest a rapid and temporally coordinated link between eye movements and language comprehension. The time course of eye movements to objects in relation to the linguistic input is crucially sensitive to referential and semantic world-language relationships, as well as to pragmatic processes. Interpretation preferences emerged in the distribution of visual attention, and between-group differences (in literacy or age) emerged in the time course and the distribution of visual attention respectively. Their sensitivity to these different factors (e.g., referential relations, pragmatic processes or between-group differences amongst others) makes eye movements a useful measure for examining visually situated language comprehension.

In more detail, we establish reference within a few hundred milliseconds (Allopenna, Magnusson, & Tanenhaus, 1998; Tanenhaus et al., 1995), a claim corroborated by the fact that listeners swiftly inspect named objects. This behavior is robust and permits researchers to gain insights into visually situated language comprehension. In the above case, looks to an object have been interpreted as reflecting that the listener is thinking about that object and has thus established reference to it. The rapid inspection of named objects also highlights that language comprehension is closely temporally coordinated with visual attention (e.g., Altmann & Kamide, 1999; Knoeferle & Crocker, 2006, 2007; Tanenhaus et al., 1995). This coordination is robust even when language refers to things that were (but no longer are) in front of our eyes (Altmann, 2004), when language is abstract (Duñabeitia, Avilés, Afonso, Scheepers, & Carreiras, 2008), and when objects are mentioned in rapid succession (Andersson, Ferreira, & Henderson, 2011). It is also present in young infants (e.g., from around 6 months of age for basic nouns, see Bergelson & Swingley, 2012; from around 36 months of age for color adjectives, Fernald, Thorpe, & Marchman, 2010).

Importantly, changes in the time course of this temporal coordination and in the distribution of visual attention reflect differences between referential relationships (*beaker* referring to a beaker) and a less-perfect fit between an object and its name (e.g., similar-sounding or rhyme words, see Allopenna et al., 1998 for a formal linking hypothesis). For *beaker*, participants inspected both the picture of a beaker and a picture of a phonological neighbor (a beetle) more often than unrelated targets from around 200 ms after word onset (e.g., see Dahan, 2010). This lasted until approximately 400 ms after which fixations to the beetle decreased. When the target word (*beaker*) rhymed with the name of another object (a speaker), then the speaker object attracted more looks from around 300 ms than phonologically unrelated objects. Thus, referents are fixated as their name unfolds and other objects are fixated to the extent that they overlap with the target in name.

It may thus be tempting to argue that a single (referential) attention mechanism serves to relate linguistic representations to representations of objects based on the goodness of the match between an object's name and potential referents. However, visual features of unmentioned objects (e.g., their color or shape) can also modulate the time course with which listeners distribute their visual attention. When participants were instructed to move a snake (depicted as stretched out) to another location, then a nearby object (a rope) depicted in a prototypical snake-shape (coiled-up) was inspected less often than the snake, but more often than unrelated objects (Dahan & Tanenhaus, 2005). These eye movements occurred approximately 200-300 ms after target word onset, and thus with a similar time course as when phonological match disambiguated reference. Semantic associations between objects also affected listeners' visual attention rapidly. 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) also attracted more attention than unrelated objects. With both a piano and trumpet present, listeners inspected the piano on the majority of trials, but the trumpet on a higher percentage of trials than unrelated objects (Huettig & Altmann, 2005, see Yee & Sedivy, 2006).

Thus, more than a referential mechanism is implicated; and just as comprehenders compute a match based on referential relations, other semantic or visual features of unnamed objects also rapidly affect the distribution of visual attention. Referents are fixated as their name unfolds and non-referents are fixated to the extent that they overlap with the target in name, semantic or visual features.

Interestingly, the time course and distribution of visual attention is also sensitive to linguistic dimensions such as the concreteness versus abstractness of a word (e.g., Duñabeitia et al., 2008). Abstract nouns and scalar quantifiers such as *some* do not have clear referents and are thus useful to consider when assessing whether a referential mechanism is sufficient to accommodate visually-situated language comprehension. When listeners heard a semantically-associated abstract word such as Spanish 'smell', they inspected the target picture (of a nose) more often and earlier (from around 200 ms after word onset) compared with their inspection of another target picture (of a baby) in response to an associated concrete word such as 'crib' (from 400 ms after word onset, association strength was controlled). Such gaze differences in the inspection of the depicted nose and the baby were absent for picture naming, in which case participants inspected referents more than semantically-associated objects (Duñabeitia et al., 2008). Thus, abstract nouns are organized primarily through semantic associations (e.g., between *smell* and the picture of a nose) while concrete nouns are organized through referential congruence (e.g., as between *nose* and the picture of a nose).

The time course of eye movements was further sensitive to the computation of scalar implicature (Huang & Snedeker, 2009). A depicted girl and a boy were each "handed" two (depicted) socks, and another girl received three balls. When the instruction was to *point to the girl that has two…,* participants' inspections to the girl with two socks rose from around 200 ms after the onset of *two*. By contrast, for the scalar quantifier *some*, inspections to the same girl rose only much later (around 1000 ms after the onset of *some*, Experiments 1 and 2 in Huang and Snedeker, 2009). The authors attributed the delay to the computation of scalar implicature, since gaze pattern suggested immediate interpretation of *some* when its sense disambiguated reference (this was the case when nine socks were evenly distributed among two boys and one girl, Experiment 3, Huang and Snedeker, 2009).

Overall, these results suggest a rapid and temporally coordinated link between eye movements and language comprehension. The fact that the time course of eye gaze is sensitive to referential and semantic relationships between language and the visual world, as well as to pragmatic processes and the interpretation of abstract language, makes eye movements a useful measure for examining visually situated language comprehension.

The studies discussed above have relied on temporal characteristics of the gaze record (relative delays indicate processing differences, e.g., Huang & Snedeker, 2009) and on the distribution of attention (e.g., indicating processing of referents vs. non-referents, e.g., Dahan & Tanenhaus, 2004). A few studies have also interpreted differences in the distribution of visual attention across objects as reflecting interpretation preferences. Preferential inspection pattern emerged, for instance, in a study by Knoeferle and Crocker (2006). When the verb (e.g., 'spy-on') in a sentence about an action between two characters was compatible with either an action performed by a non-stereotypical agent (a wizard spying) or a stereotypical agent depicted as performing a mismatching action (a detective serving food), listeners' gaze pattern revealed interpretation preferences. They preferred to anticipate the agent associated with the matching action (the wizard depicted as spying) rather than the stereotypical agent (the detective). A similar preference emerged when the choice was between the target of a recently acted-upon object and another target of a future action. In this situation, listeners rapidly inspected the target of the recent action (e.g., a candelabra that had been polished) in preference to the target of a future polishing action (e.g., polishing crystal glasses, target-condition assignment was counterbalanced, Knoeferle & Crocker, 2007). The recent-event preference replicated with real-world events (Knoeferle, Carminati, Abashidze, & Essig, 2011a) and when the within-experiment frequency of future (relative to recent) actions was increased to 75 (vs. 25) percent (Abashidze, Knoeferle, & Carminati, 2013).

In addition, eye movements have been shown to reflect qualitative differences in the interpretation between different groups of comprehenders. A comparison of skilled and less-skilled comprehenders, for instance, has revealed that these two groups can recruit verb meaning (e.g., *eat*) with the same time course for anticipating a target (e.g., a cake, Nation, Marshall, & Altmann, 2003) but differ in how often and how long they fixate that target. Less skilled (vs. skilled) comprehenders made more but shorter fixations to the target only when the verb restricted reference (e.g., *eat* requires edible objects while verbs such as *move* were less restrictive). These differences in fixation pattern were associated with a range of possible factors among them poor comprehenders' need to refresh memory traces, differences in general attention, or differences in inhibiting irrelevant information (from non-target pictures). Other (temporal) aspects of the eye-movement record have also been associated with qualitative differences in the interpretation (Mishra, Singh, Pandey, & Huettig, 2011). In a similar design as the one above, illiterates failed to anticipate the target object and only inspected it from around 300 ms after the onset of its name. By contrast, literates successfully anticipated the target object before its mention for restrictive compared with non-restrictive adjectives (e.g., a high door among other objects that did not match the adjective 'high'), suggesting they but not the illiterates develop language-derived semantic expectations.

Qualitative rather than temporal differences in the distribution of visual attention emerged in the effects of facial emotion on situated language processing. When older adults inspected a happy speaker face and subsequently listened to a sentence that described a positive event, they looked at the photograph of the positive event more than when they had inspected a negative face. Younger adults, by contrast, showed such facilitation only for negative but not for positive prime faces and sentences (Carminati & Knoeferle, 2013). Visual attention and language comprehension in older compared with younger adults thus did not differ substantially concerning the time course. Rather, differences in facial valence effects on the semantic interpretation of valenced events between the two age groups emerged in preferential eye-movement responses to positive compared with negative prime faces.

In summary, the timing and distribution of visual attention across objects and events can reflect subtle differences in comprehension and the processing of different language-"world" relationships[2]. A first important point was that visual attention is closely *temporally coordinated* with language comprehension, a link that only broke down for vague or ambiguous relationships between language and the visual world (in fact, reflecting subtle differences in comprehension when reference was ambiguous). A second important point was that most attention goes to referents for concrete words whereby semantically or visually related non-referents also attract some attention. For abstract compared with concrete words, more and earlier looks land on semantically associated objects. Thus, the distribution and time course of visual attention can index the processing of *different word-object relationships*. Further, visual attention revealed *interpretation preferences*. This was evident in comprehenders' preferred reliance during comprehension on a recent action target over anticipating a future action target. Finally, eye gaze measures revealed *qualitative* differences in comprehension between groups of comprehenders.

However, we know preciously little about which linguistic, cognitive or social factors affect which aspect of the eye-movement record. Differences in

[2] For other measures and linking hypotheses, see also Altmann, 2010; Arai, Van Gompel, & Scheepers, 2007; Engelhardt, Ferreira, & Patsenko, 2010; Scheepers & Crocker, 2004.

comprehension skill, for instance, affected the duration and frequency of eye gaze but not its time course. Age-related differences in emotion interpretation also emerged in the duration and frequency of attention rather than time course differences. By contrast, differences in literacy affected the time course of object-directed gaze.

2.2 Linking issues in eye-tracking studies: semantic versus syntactic processes

All in all, there is considerable ambiguity as to which specific comprehension (sub)-processes are reflected at any given point in time by the single stream of eye movements that we record. Since the linking relies predominantly on properties of the design (minimal comparisons between individual conditions), and relative timing of the eye movements rather than on distinct fixation signatures, even minor weaknesses in the design can lead to ambiguity in linking eye movements to cognitive processes. Consider a study by Altmann and Kamide (1999) which formed the basis for the above-reviewed research on qualitative differences (e.g., Nation et al., 2003). Altmann and Kamide examined eye movements to a target object (a cake) for selective verbs such as *eat* compared with non-selective verbs such as *move*. The picture showed a boy, a cake, and three (inedible) toys. The expectation was that the selectional restrictions of the verb *eat* in *The boy will eat…* would guide the listeners' attention to the one edible object (a cake) before its mention. For the non-restrictive verb *move*, by contrast, listeners' should distribute their visual attention evenly across the four objects since all of them were moveable. Earlier eye movements to the cake for *eat* than for *move* verbs were taken to reflect the differential in verb selection restrictions. However, instead of verb selection restrictions, semantic associations between *eat* and the depicted cake could have triggered these eye movements, since the design did not control for this possibility and since we do not know whether verb selection restrictions and semantic associations are associated with distinct fixation signatures.

One solution to this problem has been to improve the design (e.g., Kamide, Altmann, & Haywood, 2003). An additional means has been to rely on complementary measures to reduce ambiguity in the linking hypothesis of a single measure (see also Willems, Özyürek, & Hagoort, 2008; Knoeferle et al., 2011 for related approaches with different measures). An example comes from a pair of studies that examined visual context effects on the disambiguation of local structural ambiguity using eye tracking and ERPs. In the initial eye-tracking study, Knoeferle et al. (2005) examined the processing of German sentences with a sentence-initial structural ambiguity which was disambiguated by case marking on the determiner of a sentence-final noun phrase (e.g., *Die Prinzessin malt der Fechter*, 'The princess (object, ambiguous) paints the fencer (subject)'). Earlier disambiguation was possible if comprehenders related the verb to depicted thematic relations between the referents of the two noun phrases (the princess was depicted as washing a pirate and a fencer was depicted as painting the princess). As comprehenders heard 'The princess (amb.) paints…' they rapidly related the verb 'paints' to the action of the fencer and anticipated the fencer before it was mentioned. The visual anticipation of the fencer was interpreted as reflecting assignment of an agent role to 'fencer' and of a patient role to 'princess', indicating disambiguation of the syntactic and thematic role relations before linguistic disambiguation through case marking on 'the' (*der*) in 'the fencer' (*der Fechter* is in nominative case and is marked as the subject and agent of the sentence).

Strictly speaking, however, this gaze pattern (eye movements to the fencer before its mention) could also index an initial lexical mismatch between the action of the princess (washing) and the verb ('paints'). Upon noticing that 'paints' mismatches the action of the princess, comprehenders begin to search for a matching instrument and this leads them to the action of the fencer and to increased inspection of the associated fencer. Again, since we do not know whether particular eye-movement signatures correspond to specific comprehension sub-processes such as referential matching and structural disambiguation (to the extent that such one-to-one linking exists at all), unambiguous interpretation of these fixation patterns is difficult. To reduce the ambiguity in the linking of visual attention to comprehension (or other cognitive) sub-processes, an ensuing study complemented the eye movements with another continuous measure (event-related brain potentials, of which more below).

2.3 Visually situated language comprehension: evidence from ERPs

Event-related brain potentials are a useful complementary measure since they reflect cognitive processes over time and vary both temporally and qualitatively (in their polarity) for lexical-semantic compared with compositional and syntactic processes. Lexical-semantic processes and the integration of new meaning in the semantic context have been associated with the so-called N400 effect. The effect is a negative deviation in mean amplitude ERPs approximately 400 ms after an event such as the presentation of a word or picture. The better a word fits into the preceding context, the more the mean amplitude N400s decrease (e.g., Kutas & Hillyard, 1980, 1984; Kutas & Federmeier, 2011). By contrast, syntactic revision and the processing of syntactic violations have been associated with a qualitatively distinct effect, the so-called P600 (also called syntactic positive shift, e.g., Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992, 1993). This is a positive deviation in mean amplitude ERPs approximately 600 ms after a stimulus[3].

Complementing eye movement studies with ERP studies can help us discard alternative interpretations of the data from an individual measure. Knoeferle, Habets, Crocker, and Münte (2008) did just that with the materials from the eye tracking study by Knoeferle et al. (2005), and recorded event-related brain potentials as participants inspected similar event depictions and listened to related German sentences with an initial structural and role ambiguity. When the verb related to an event depicting the referent of the first noun phrase as the patient (vs. agent), mean amplitude ERPs to the verb were more positive (P600). In an audio-only baseline condition, these mean amplitude P600 differences emerged only later, at a point in time when case marking on the determiner of the second noun phrase disambiguated towards the object-subject order. Given the comparison of the audio-visual condition with the audio-only baseline, and given the interpretation of the P600 as an index of structural revision and structural disambiguation, the depicted events likely triggered revision processes and not just a lexical-semantic mismatch and ensuing visual search for a matching action.

In this particular case, the eye-movement patterns together with the P600 suggested that the underlying processes involved the anticipation of role fillers and

[3] Note that the distinction between the N400 and P600s is not entirely clear-cut, as a ‘semantic’ P600 emerged in response to what looked like semantic violations (Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). Naturally, ambiguity in the linking hypotheses leads to ambiguity in understanding and modeling language comprehension processes. Interpretation problems resulting from ambiguity have also been discussed elsewhere (see Kutas, Van Petten, & Kluender, 2006 for ERPs; Tanenhaus, 2004 for eye tracking) and one proposal for eye-movement data has been to more explicitly and formally specify one’s linking hypotheses (e.g., Allopenna et al., 1998; Tanenhaus, 2004).

associated structural revision. A similar distinction between the P600 and the N400 (reflecting lexical-semantic processes) is also apparent in research on the integration of co-speech gestures (Holle et al., 2012; Kelly, Kravitz, & Hopkins, 2004).

Different meaning relationships between a gesture and speech, for instance, elicited distinct N400 effects but no P600 differences (Kelly et al., 2004). Comprehenders inspected a gesture that related to speech in three ways. The gesture matched speech (it underscored the verbally-expressed thinness of a glass), was complementary to speech (it related to the thinness of the glass while the speech mentioned tallness of the glass), or contradicted the speech (the gesture described the shortness of a dish while a tall glass was mentioned). A fourth, speech-only / no-gesture condition served as a baseline. The condition without gestures elicited larger broadly distributed N400 effects relative to the other three conditions. In addition, larger mean amplitude N400s emerged in the no-gesture than the other three conditions over anterior sites. Over bilateral temporal sites, ERPs to the gesture mismatches were crucially more negative than ERPs to the matches (but not to the other conditions). Gesture mismatches differed from ERPs to the complementary and the no-gesture conditions in particular over the right hemisphere (Kelly et al., 2004; see also Wu & Coulson, 2005, 2007 for relevant results, Kelly, Creigh, & Bartolotti, 2009 on the automaticity of such integration and Kelly & Breckinridge Church, 1998 on developmental differences).

The semantic interpretation of co-speech gestures crucially depends on the relative timing of speech and co-speech gestures. When the gestures were presented together with speech (zero ms delay) or when speech was delayed by 160 ms relative to the onset of the corresponding gesture, mean amplitude N400s were larger for gesture-speech mismatches than matches. This N400 difference was absent when the gesture preceded the corresponding word by 360 ms (Habets, Kita, Shao, Özyurek, & Hagoort, 2011). In sum, variation in the N400 mean amplitudes indexed subtle differences in the semantic contribution of iconic gestures to the interpretation, and temporal coordination seems to be a key factor in the successful integration of language and non-verbal cues (see also section 2.1).

Another kind of gesture (beat gestures) can affect comprehension processes such as structural disambiguation in the face of temporary linguistic ambiguity (subject-object, SO compared with object-subject, OS, Holle et al., 2012). A verb following the ambiguous noun phrase sequence (subject-object or object-subject) resolved the temporary ambiguity towards either the SOV or OSV structure. In addition to linguistic disambiguation, a video-taped speaker emphasized either none of the constituents, the first, or the second noun phrase. Control conditions showed a red dot moving along the gesture trajectory. Analyses of participants' accuracy scores on 'yes/'no' questions about thematic role relations revealed lower accuracy for object than subject-initial sentences.

In the ERPs, object compared with subject-initial sentences elicited an anterior negativity followed by a relative posterior positivity to the disambiguating verb in the absence of beat gestures. This relative P600 to the disambiguating verb remained virtually unchanged when a beat gesture emphasized the first noun phrase; by contrast, a beat gesture on the second, ambiguous noun phrase eliminated the P600 and only the anterior negativity remained (Experiment 1). Since the P600 difference in response to structural disambiguation at the verb was eliminated neither by an auditory pitch accent (Experiment 2) nor by the moving red dot (Experiment 3), the authors concluded that the beat gesture affected syntactic structuring. The beat could highlight relevant information for a short period of time, which would explain the absence of gesture effects when it occurred long before disambiguation (at the first noun phrase). Alternatively, or in addition, a beat signals the sentential subject, in which case a beat on the first noun phrase is redundant since the first noun phrase is assumed to be the subject. A beat on the second of two noun phrases, by contrast, signals that the subject is in an unusual position and thus likely has a disambiguating effect.

Thus, qualitatively distinct ERP congruence effects indexed whether visual context (gesture and action events) influenced semantic interpretation (as indexed by N400 mean amplitude differences) or rather structural disambiguation (as indexed by P600 mean amplitude differences), a distinction which was useful for disambiguating alternative accounts of eye-movement results.

2.3.1 Linguistic versus visual context effects: same or different ERP effects?

Section 2.3 has argued that complementing eye movement studies with ERP studies can help us discard alternative interpretations of the data from an individual measure. The qualitative distinction between the N400 and the P600 was indeed used to this effect, with results suggesting rapid effects of non-linguistic cues on both semantic interpretation and syntactic disambiguation. Given that non-linguistic cues rapidly affect these comprehension processes, one might assess whether their effects on comprehension are qualitatively similar to those of linguistic cues. Finding non-linguistic cues on a par with linguistic ones would be a strong argument in favor of examining language comprehension in rich non-linguistic contexts.

Research on the semantic processing of pictures, for instance, suggests that semantic matching of pictures elicits negativities which differ at least partially from the negativities elicited by semantic interpretation in strictly linguistic contexts. One is an earlier anterior N300 difference (larger to picture-picture mismatches than matches), and the other is a later posterior negativity likened to the semantic N400 in verbal stimuli (e.g., Barrett & Rugg, 1990). The interpretation of pictures (compared with words) in sentence contexts also seems to elicit a different topography in the elicited N400 effect (pictures: anterior; words: posterior, Ganis, Kutas, & Sereno, 1996). In addition adjectival color mismatches in the token test (object: red square; linguistic input: green square) yielded an anterior N2b component instead of a posterior N400. The N2b has been associated with mismatch detection rather than language processing (D'Arcy & Connolly, 1999; Vissers et al., 2008).

By contrast, other picture-sentence congruence manipulations (e.g., noun-object: Friedrich & Friederici, 2004; verb-action, Knoeferle et al., 2011b) yielded N400s akin to those in language comprehension tasks and no N2b differences (e.g., Kutas, 1993; see also Kutas, Van Petten, & Kluender, 2006; Otten & van Berkum, 2007; Van Berkum et al., 1999), suggesting visual contexts modulated language comprehension and not other processes. Likewise, when depicted agent-action-patient events enabled structural disambiguation (Knoeferle et al., 2008), the topography of P600 differences was visually indistinguishable relative to cases when structural disambiguation was enabled by case marking on the determiner of a noun phrase. Further evidence for similarities in how linguistic compared with pictorial cues affect comprehension comes from a study by Willems, Özüyrek, and Hagoort (2008). They examined the time course of the semantic integration of a word and picture with a previous sentence context. Sentences either contained no mismatch, a mismatching word, a mismatching picture, or both picture and word mismatches. In the ERPs, they observed an N400 effect which was similar for pictures and words in terms of latency, topography, and amplitude and no clear evidence for a picture-specific N300 (see also Ganis et al., 1996).

Thus, for word-object mismatches, and for pictorial stimuli, ERP patterns in response to incongruence seem be at least partially distinct compared with semantic interpretation in strictly linguistic contexts. However, for real-time sentence interpretation in rich contexts, visual context effects on both semantic and syntactic processes resembled the ERP effects observed for these processes in purely linguistic contexts (although the presence of visual information can shift the distribution of the N400). Thus, it seems that sentence comprehension draws on linguistic and non-linguistic information with the same time course and also recruits at least partially overlapping brain areas (see Willems et al., 2008).

2.3.2 Ambiguity in linking of ERP effects to comprehension sub-processes

While these experiments – by virtue of design constraints – interpreted the effects of visual context on comprehension as either semantic or syntactic, not all distinct sentence-picture relationships are dissociable by means of ERPs. Vissers et al. (2008), for instance, observed statistically indistinguishable P600 differences in response to two kinds of spatial mismatches (vs. matches). When participants verified depictions (□ △) against ensuing written sentences (e.g., ‘the triangle stands behind / in front of / above the square’), the two mismatches (‘in front of / above’) elicited statistically indistinguishable mean amplitude negativities and P600s relative to the matches (‘behind’).

Vissers et al. (2008) argued that the absence of a difference between distinct picture-sentence mismatches reflects a general monitoring mechanism responding to any kind of violation. If that is true, however, then it is unclear why picture-sentence incongruence processing does not always elicit a P600. When participants inspected scenes depicting an agent-action-patient event and subsequently read a related sentence in which the verb either matched or mismatched the previously inspected action, no P600 differences emerged (Knoeferle et al., 2011b). Instead, mean amplitude N400s to the verb were more negative and post-sentence verification response times were longer for verb-action mismatches than matches.

Further support for the view that ERPs are sensitive to distinct picture-sentence relations comes from two studies on thematic role assignment. In one study, healthy older adults verified whether a spoken Dutch sentence (e.g., 'The tall man on [sic] this picture pushes the young woman') accurately described a previously-inspected line drawing (e.g., a man pushing a woman versus a woman pushing a man, Wassenaar & Hagoort, 2007). For active sentences, a reliably larger posterior negativity was followed by a numerically larger positivity to role relation mismatches (vs. matches) at the verb (centro-posterior from 50-450 ms; for anterior sites from ca. 50-300 ms), and by a broad negative shift to mismatches relative to matches in the post-verbal noun. ERPs to irreversible active and reversible passive sentences showed an early negativity, a subsequent late positivity, and a negative shift to mismatches (vs. matches). These effects were interpreted as broadly reflecting thematic role assignment with subtle differences depending on sentence type.

In another study, thematic role assignment effects were differentiated by ERPs from lexical verb-action mismatches (Knoeferle, Urbach, & Kutas, 2010, under revision). Participants read a subject-verb-object sentence (500 ms SOA in Experiment 1), and verified post-sentence whether or not the verb and/or the thematic role relations matched a preceding picture (depicting two participants engaged in an action). Sentences either matched the picture, or mismatched in either the action or the depicted role relations, or both. These two types of mismatches (actions vs. role relations) yielded different ERP effects. Role-relation mismatch effects emerged as anterior negativities to the mismatching subject noun, and preceded action mismatch effects (centro-parietal N400s greater to the mismatching verb).

Overall, thus, more than a single mechanism is active in picture-sentence congruence processing. Distinct picture-sentence mismatches do elicit distinct ERP patterns and can distinguish lexical-semantic from compositional thematic effects. Perhaps the null effect in Vissers et al. is genuine (people do not differentiate the different spatial relations online) but they do differentiate between other picture-sentence relations. Alternatively, the chosen measure (ERPs) was insensitive to the difference. Clearly, much remains to be learned about the nature of the relationship between ERPs (or eye movements) and visually situated language processing.

3. Summary and conclusions

A first section (1.1) documented that theories providing detailed accounts of sentence processing have focused on accommodating comprehension in strictly linguistic contexts. By contrast, we ultimately want to accommodate comprehension in all sorts of situations, including those that feature a rich non-linguistic context. Section 1.2 proceeded to outline how methodological advances and the combination of continuous measures with visually situated tasks has facilitated research on real-time visually situated language comprehension. Section 2 characterized visually situated language comprehension by reviewing both eye-tracking and ERP evidence. In addition, it highlighted ambiguity in the linking assumptions where relevant.

What we can take away from this review is that referential processes are central in visually situated spoken language comprehension but that eye movements are also exquisitely sensitive to other aspects of language or the visual context. Among these aspects are the abstractness (vs. concreteness) of language, vagueness and scalar implicature, word order, and action events. When world-language relations are underspecified or when sentences are difficult, object-based eye gaze is delayed, suggesting it is sensitive to important aspects of the comprehension process. Eye movements have, in addition, revealed interpretation preferences of depicted events over, for instance, the anticipation of future events. In addition, they have been interpreted as reflecting qualitative differences in the comprehension and attention processes of illiterates and literates and between high and low skill comprehenders.

However, weaknesses in the design complicate the unambiguous interpretation of the gaze pattern, suggesting we must strive to improve our linking hypotheses. One way to do this, as outlined in section 2, is to complement eye-tracking studies with ERP studies to narrow the interpretation of the gaze pattern. Indeed, ERPs offer a relatively robust distinction between semantic and syntactic processes and can help us ground the interpretation of the eye-tracking data. This became clear when reviewing the effects of action events on structural disambiguation and the effects of co-speech gestures on semantic interpretation (e.g., distinct for complementary vs. mismatching gesture-sentence pairs). Distinct effects of (beat) gestures on structural disambiguation corroborated the usefulness of the N400-P600 distinction for the investigation of visually situated language comprehension. Visual and linguistic cues seemed furthermore on a par in informing sentence comprehension, as revealed by highly similar ERPs independent of whether linguistic or non-linguistic cues contributed to the interpretation and to structural disambiguation in rich contexts.

And yet, one wonders whether long-term, ERPs alone are sufficient as a window into (potentially subtle) effects of how pictorial information contributes to language comprehension. Recall that in the study by Vissers et al. (2008) mismatches between the spatial configuration of objects and different prepositions ('in front of' relative to 'above') did not elicit a different ERP pattern. It's possible that this null effect reflects a genuine absence of differences, as argued by the authors. Imagine, however, that we presented sentences such as 'The triangle stands in front of / above the square' together with the depiction of several objects, among them a triangle to the right of a square, and tracked listeners' eye movements. At 'triangle', listeners would inspect the triangle, and at 'stands in front of, we may expect them to anticipate the location indicated by compositional interpretation of the noun, the verb, and the preposition (to the right of the triangle, see Burigo & Knoeferle, 2011; Chambers et al., 2002 for evidence of anticipatory eye movements following spatial prepositions). By contrast, for 'The triangle stands above the square, listeners should anticipate the location below the triangle (since the triangle is above that location). Thus, eye-movement behavior in this kind of study would reveal a distinct distribution of visual attention for these two mismatches. This in turn could influence which other information is perceived and is available for comprehension.

Admittedly, the paradigm envisaged in this example (visual inspection of objects during spoken comprehension) differs from the one used by Vissers and colleagues (pictures followed by written sentences) and this may change the integration of language and the visual context. However, the example illustrates the potentially enriching effect of combining eye tracking with EEG recordings (across studies and within the same experiment). This approach could constrain the interpretation of individual measures. It could also pave the way for extending current accounts of visually situated language comprehension (e.g., Knoeferle & Crocker, 2006, 2007), which have largely been shaped by eye-movement results, with a description of the functional brain correlates implicated in visually situated language processing.

Indeed, the first step in this direction has been undertaken using a connectionist model of visually situated language comprehension (Crocker, Knoeferle, &

Mayberry, 2010). The model's task is to predict thematic roles fillers in a target output representation during sentence processing. One type of sentences tested is the structurally ambiguous German subject-verb-object and object-verb-subject sentences from Knoeferle et al. (2005, see Section 2). As reviewed in Section 2, human participants anticipate role fillers visually when the verb identifies the action in a depicted agent-action-patient event. The model predicted the correct role-fillers at the same point in time as comprehenders' gaze reflected the anticipation of the correct role filler. In the ERPs, comprehenders exhibited a P600, larger to object- than subject-initial sentences time-locked to the onset of the verb and leading into the post-verbal region. In the model, effects of structural revision (through linguistic cues and depicted events) were examined through changes of hidden-layer activation from the processing step at the verb to the next word (i.e., after event-based disambiguation). These changes were larger for object-initial sentences compared to subject-initial sentences, suggesting structural revision.

In sum, such a cross-methodological venture has the potential to enrich extant models of visually situated language comprehension with measures that permit us to monitor comprehension from stimulus presentation to an overt visual response. In addition, it can enrich our understanding of how visual attention and brain responses are related to different aspects of visually-situated language comprehension, thus permitting us to refine our interpretation of individual measures.

References

Abashidze, D., Knoeferle, P., & Carminati, M. N. (2013). Do comprehenders prefer to rely on recent events even when future events are more likely to be mentioned? In: *Proceedings of the Conference on Architectures and Mechanisms for Language Processing*. Marseille, France.

Allopenna, P., 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.

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

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

Altmann, G.T.M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33,* 583-609.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. Cognition, 30, 191–238.

Andersson, R., Ferreira, F., & Henderson, J. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica, 137,* 208-216.

Barrett, S. E., & Rugg, M.D. (1990). Event-related brain potentials and the matching of pictures. *Brain and Cognition, 14,* 201-212.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA, 109, 3253-3258.*

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113,* 787-821.

Burigo, M., & Knoeferle, P. (2011). Visual attention during spatial language comprehension: Is a referential linking hypothesis enough? In: Carlson L, Hölscher C, Shipley T (Eds). *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, Tx: Cognitive Science Society.

Carminati, M. N. & Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PLoS ONE, 8(9): e72559. doi:10.1371/journal.pone.0072559*

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A Psycholinguistic Processing Model of Verification. *Psychological Review, 82*, 45–73.

Chambers, C. G., Tanenhaus M.K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30-49.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht: Kluwer.

Crocker, M.W., & Brants. T. (2000). Wide Coverage Probabilistic Sentence Processing. *Journal of Psycholinguistic Research, 29*, 647-669.

Crocker, M. W., Knoeferle, P., & Mayberry, M. (2010). Situated sentence comprehension: The coordinated interplay account and a neurobehavioral model. *Brain and Language, 112*, 189–201.

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. (2010). The time course of interpretation in speech comprehension. *Current Directions in Psychological Science, 19*, 121-126.

D'Arcy, R. C. N., & Connolly, J. F. (1999). An event-related brain potential study of receptive speech comprehension using a modified Token Test. *Neuropsychologia, 37*, 1477–1489.

Duñabeitia, J. A., Avilés, 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.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology, 60*, 190–217

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Waler (Eds.), Sentence processing: psycholinguistic studies presented to Merrill Garrett (pp. 27–85). Hillsdale, NJ: Lawrence Erlbaum.

Frazier, L., & Fodor, J. D. (1979). The sausage machine: a new two-stage parsing model. *Cognition, 6*, 291–325.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6,* 78-84.

Friedman, A. & Bourne Jr., L. E. (1976). Encoding the levels of information in pictures and words. *Journal of Experimental Psychology: General, 105*, 169-190.

Ganis, G., Kutas, M., & Sereno, M. I. (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.

Garrod, S. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition 27*, 181–218.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gorrell, P. (1995). Syntax and parsing. Cambridge: Cambridge University Press.

Gough, P. B. (1965). Grammatical transformations and speed of understanding. *Journal of Verbal Learning & Verbal Behavior, 4,* 107–111.

Habets, B., Kita, S., Shao, Z., Özyürek, A. & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*, 1845-54.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439-483.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*, 883–899.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32*, 101–122.

Holle, H, Obermeier, C., Schmidt-Kassow, M., Friederici, A.D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology 3*, 1-12.

Huang, Y. T., & Snedeker, J. (2009) Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology, 58,* 376-415.

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,* B23-32.

Johnson-Laird, P. (1981). Comprehension as the construction of mental models. *Philosophical Transactions of the Royal Society, Series B, 295*, 353-374.

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time course of prediction in incremental sentence processing: evidence from anticipatory eye-movements. *Journal of Memory and Language, 49,* 133-156.

Kelly, S. D., & Breckinridge Church, R. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69,* 85-93.

Kelly, Spencer D., Creigh, P., & Bartolotti, J. (2009). Integrating speech and iconic gestures in a stroop-like task: evidence for automatic processing. *Journal of Cognitive Neuroscience 22*. 683-94.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language 89*. 253–60.

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.

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, 75*, 519–543.

Knoeferle, P., Habets, B., Crocker, M.W., & Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cerebral Cortex 18*. 789-95.

Knoeferle, P., Urbach, T., & Kutas, M. (2010). Verb-action versus role relations congruence effects: Evidence from ERPs in sentence-picture verification. In S. Ohlsson and R. Catrambone (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (p.2446-2451). Austin, TX: Cognitive Science Society.

Knoeferle, P., Carminati, M. N., Abashidze, D., & Essig, K. (2011a). Preferential inspection of recent real-world events over future events: evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology, 2*, 376. doi:10.3389/fpsyg.2011.00376

Knoeferle, P., Urbach, T, & Kutas, M (2011b). Comprehending how visual context influences incremental sentence processing: insights from ERPs and picture-sentence verification. *Psychophysiology, 48*, 495-506.

Kolk, H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language, 85,* 1-36.

Kuperberg G. R., Sitnikova T., Caplan D., Holcomb P.J. (2003) Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 217*, 117-129.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes, 8*, 533-572.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161– 163.

Kutas, M., & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62,* 621-647.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II. In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of Psycholinguistics* (2nd edition, pp. 659–724). New York: Elsevier Press.

Roger Levy. 2008. Expectation-based syntactic comprehension. *Cognition 106*, 1126-1177.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Mayberry, M., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science, 33*, 449-496.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2011). Spoken language-mediated anticipatory eye- movements are modulated by reading ability - Evidence from Indian low and high literates. *Journal of Eye Movement Research, 5,* 1-10.

Mitchell, D. C., Cuetos, F., Corley, M., & Brysbaert, M. (1995). Exposure-based models of human parsing: evidence for the use of coarse-grained (nonlexical) statistical records*. Journal of Psycholinguistic Research, 24*, 469-488.

Nation, K., Marshall, C., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86,* 314-329.

Novick, J., Thompson-Schill, S., & Trueswell, J. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition, 107,* 850-903.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785-806.

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and sce- nario fit on the discourse-dependent N400 effect. *Brain Research, 1153*, 166-177.

Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169-190.

Potter, M. C, Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General, 115*, 281-294.

Tanenhaus, M. K., Carroll, J. M., & Bever, T. G. (1976). Sentence- picture verification models as theories of sentence comprehension: A critique of Carpenter and Just. *Psychological Review, 83*, 310-317.

Tanenhaus, M.K. (2004). On-line sentence processing: past, present and, future. In M. Carreiras and C. Clifton, Jr. (eds). *On-line sentence processing: ERPS, eye movements and beyond*. Psychology Press, pp. 371-392.

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.

Trueswell, J.C. & Tanenhaus, M.K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In: Clifton, Frazier & Rayner (Eds.). Perspectives on Sentence Processing, 155-179. Hillsdale, NJ: LEA Press.

Spivey-Knowlton, M., & Tanenhaus, M. (1998). Syntactic Ambiguity Resolution in Discourse: Modeling the Effects of Referential Context and Lexical Frequency. *Journal of Experimental Psychology: LMC*, 24, 1521-1543.

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

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241–255.

Vissers, C., Kolk, H., Van de Meerendonk, N., & Chwilla, D. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia, 46*, 967–982.

Wassenaar, M., & Hagoort, P. (2007). Thematic role assignment in patients with Broca's aphasia: Sentence-picture matching electrified. *Neuropsychologia, 45,* 716–740.

Willems, R. M., Ozyurek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience, 20,* 1235-1249.

Wu, Y.C. & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. Psychophysiology 42: 654-667.

Wu, Y.C. & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain & Language 101*, 234-245.

Yee, E., & Sedivy, J. (2006). Eye Movements to Pictures Reveal Transient Semantic Activation During Spoken Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition 32*, 1–14.