

Visual Scenes Trigger Immediate Syntactic Reanalysis: Evidence from ERPs during Situated Spoken Comprehension

Pia Knoeferle¹, Boukje Habets², Matthew W. Crocker³ and Thomas F. Münte²

¹Center for Research in Language, University of California San Diego, USA, ²Department of Neuropsychology, Otto-von-Guericke University Magdeburg, Germany and ³Department of Computational Linguistics, Saarland University, Germany

A central topic in sentence comprehension research is the kinds of information and mechanisms involved in resolving temporary ambiguity regarding the syntactic structure of a sentence. Gaze patterns in scenes during spoken sentence comprehension have provided strong evidence that visual scenes trigger rapid syntactic reanalysis. However, they have also been interpreted as reflecting nonlinguistic, visual processes. Furthermore, little is known as to whether similar processes of syntactic revision are triggered by linguistic versus scene cues. To better understand how scenes influence comprehension and its time course, we recorded event-related potentials (ERPs) during the comprehension of spoken sentences that relate to depicted events. Prior electrophysiological research has observed a P600 when structural disambiguation toward a noncanonical structure occurred during reading and in the absence of scenes. We observed an ERP component with a similar latency, polarity, and distribution when depicted events disambiguated toward a noncanonical structure. The distributional similarities further suggest that scenes are on a par with linguistic contexts in triggering syntactic revision. Our findings confirm the interpretation of previous eye movement studies and highlight the benefits of combining ERP and eye-tracking measures to ascertain the neuronal processes enabled by, and the locus of attention in, visual contexts.

Keywords: cross-methodological approach, depicted events, event-related potentials, situated language comprehension

Introduction

The monitoring of eye movements to objects in visual scenes during spoken sentence comprehension has provided strong evidence for the view that scene information can rapidly influence incremental sentence comprehension (e.g., Tanenhaus et al. 1995; Sedivy et al. 1999; Chambers et al. 2004; Knoeferle et al. 2005; Knoeferle and Crocker 2006, forthcoming). For instance, it has been shown that the type of visual referential context influences the initial structuring and interpretation of temporarily ambiguous instructions such as “put the apple on the towel in the box” (Tanenhaus et al. 1995). The temporary ambiguity results from the fact that 2 alternative ways of structuring and interpreting the sentence fragment are possible: the phrase “on the towel” can either be attached to the noun phrase “the apple,” indicating the location of the apple, or be attached to the verb (*put*) and interpreted as a destination. When scenes contained one apple on a towel and an empty towel (representing a destination), gaze patterns to the empty towel revealed that people rapidly interpreted the ambiguous phrase “on the towel” as destination. The interpretation revealed in the gaze pattern was taken as an indication of the structure that people built, attaching the prepositional phrase to the verb. In contrast, when scenes contained 2 apples only

one of which was on a towel, the absence of looks to the empty towel (the destination) suggested that the phrase “on the towel” had been interpreted as the location of the apple and attached to the first noun phrase.

These findings on the rapid use of scene information for the incremental syntactic structuring of an utterance extend to scenes that contain depicted agent-action-patient events (Knoeferle et al. 2005). While inspecting such scenes, participants heard related German sentences that had canonical (subject-verb-object) or noncanonical (object-verb-subject) word order. Both of these orders are grammatical in German, but the subject-first order is preferred (e.g., Matzke et al. 2002; see also Schlesewsky et al. 2000). Just as in the studies by Tanenhaus et al. (1995), sentences in the studies by Knoeferle et al. (2005) contained a temporary structural ambiguity: the first noun phrase could initially be interpreted as the subject (agent) or object (patient) of the sentence, temporarily allowing both a subject-verb-object and an alternative object-verb-subject structure. Utterance-based disambiguation of the first noun phrase occurred late, when linguistic cues on the determiner of the sentence-final noun phrase marked that noun phrase as either an object or a subject, resolving the ambiguity toward a subject-verb-object or object-verb-subject structure, respectively.

Scene-based disambiguation of the structural ambiguity, in contrast, was possible earlier, when the verb mediated one of 2 scene events and its depicted role relations. For canonical sentences, the verb mediated an event that depicted the referent of the initially ambiguous noun phrase (e.g., *The princess*) as the agent (princess-paints-fencer), whereas for noncanonical sentences, the verb mediated an event that depicted the princess as the patient (pirate-washes-princess) of an event. Eye movements to the fencer and the pirate shortly after the verb were interpreted as reflecting incremental assignment of a thematic role to the role-ambiguous first noun phrase. More eye movements to the fencer (the patient of the princess painting event) for structurally ambiguous canonical than noncanonical sentences and more inspections to the pirate (the agent of the pirate-washes-princess event) for noncanonical than canonical sentences, respectively, indicated that people rapidly used the depicted events for assigning the appropriate thematic role to the initially ambiguous noun phrase, suggesting they incrementally resolved the structural ambiguity.

Although these findings (Tanenhaus et al. 1995; Knoeferle et al. 2005) provide behavioral evidence for the claim that scene information affects the incremental structuring of initially structurally ambiguous utterances, utterance-mediated attention in scenes is also known to reflect various other underlying linguistic and nonlinguistic processes such as semantic interpretation

(Sedivy et al. 1999), thematic interpretation (e.g., Altmann and Kamide 1999), or visual search (e.g., Spivey et al. 2001). Eye movement measures alone furthermore do not clarify whether the processes involved in resolving local structural ambiguity through scene information are similar to those triggered when linguistic cues resolve a temporary structural ambiguity.

To better understand the influence of visual contexts (depicted events) on structural revision during spoken sentence comprehension, 2 event-related potential (ERP) studies were conducted. Measures such as ERPs have in the past been used to examine the processing of syntactic violations (e.g., Friederici et al. 1993; Hagoort et al. 1993; Osterhout et al. 1994) and, in particular, the resolution of temporary structural ambiguity through linguistic cues: when linguistic cues triggered structural revision toward a noncanonical structure during reading in the absence of scenes, the difficulty of this revision has typically been associated with a positivity that has a maximum at approximately 600 ms (P600, e.g., beim Graben et al. 2000; Frisch et al. 2002; Matzke et al. 2002).

We rely on these findings for investigating the structural revision of locally structurally ambiguous German utterances through linguistic cues (e.g., case marking on the determiner of a noun phrase) and through verb-mediated depicted events. When no scenes are present, and disambiguation of local structural ambiguity toward either an object-verb-subject or a subject-verb-object order can only occur through a case-marked determiner on the second noun phrase, we expect to replicate previous findings (e.g., Matzke et al. 2002): we should see a positivity with a peak at approximately 600 ms time locked to the onset of the second noun phrase in response to linguistic disambiguation toward the noncanonical structure.

In contrast, when scenes are present, disambiguation may occur prior to the second noun phrase at the verb: if the interpretation of the eye movement behavior for the studies by Knoeferle et al. (2005) as reflecting structural revision through verb-mediated depicted events is correct, then we should find a P600 for initially ambiguous noncanonical relative to canonical sentences time locked to the onset of the verb that identifies relevant events. For structurally unambiguous controls, or in the absence of scenes, we should find no P600 time locked to verb onset for noncanonical relative to canonical sentences.

The present study furthermore offers the opportunity to compare the neural correlates of disambiguation when it is triggered by depicted events compared with when disambiguation is enabled through linguistic cues such as case marking on the determiner of a noun phrase. Observing, for instance, components that differ in latency and/or scalp distribution in response to scene-based disambiguation compared with disambiguation through linguistic cues would suggest differences in the neural processes underlying these 2 ways of disambiguating the utterance. Alternatively, finding no clear difference when comparing the latency and topography of components in response to scene-based versus utterance-based disambiguation would support the view that similar neural processes underlie these 2 ways of structural disambiguation.

We examined these expectations by recording ERPs while people listened to initially structurally ambiguous canonical and noncanonical German sentences and to unambiguous controls in the presence of depicted event scenes (Experiment 1; audiovisual experiment). Experiment 2 examined comprehension of the same utterances in the absence of scenes (auditory experiment).

Methods

Participants

There were 16 participants in Experiment 1 and 16 participants in Experiment 2, all of whom were native speakers of German and students of the University of Magdeburg. All participants were right handed, had normal or corrected-to-normal vision and hearing, and had given written informed consent prior to the experiment.

Design and Materials

The materials derive from the stimuli of Experiment 1 in Knoeferle et al. (2005). We first describe the experimental conditions and subsequently detail the counterbalancing and material creation. Figure 1A was presented with initially ambiguous canonical (subject-verb-object, (1a)) and noncanonical (object-verb-subject, (1b)) sentences: the first noun phrase was temporarily ambiguous and could either be the subject (1a) or the object (1b) of the sentence.

- (1a) Die Prinzessin (ambiguous) malt offensichtlich den Fechter (object) (canonical),
 “the princess (amb.) paints apparently the fencer (object).”
 (1b) Die Prinzessin (amb.) wäscht offensichtlich der Pirat (subject) (noncanonical),
 “the princess (amb.) washes apparently the pirate (subject).”

In addition, unambiguous canonical and noncanonical sentences were presented with Figure 1B. We created the sentences for the unambiguous conditions by replacing the ambiguous first noun phrase (*Die Prinzessin*, “the princess,” (1a/b)) with a masculine noun phrase.

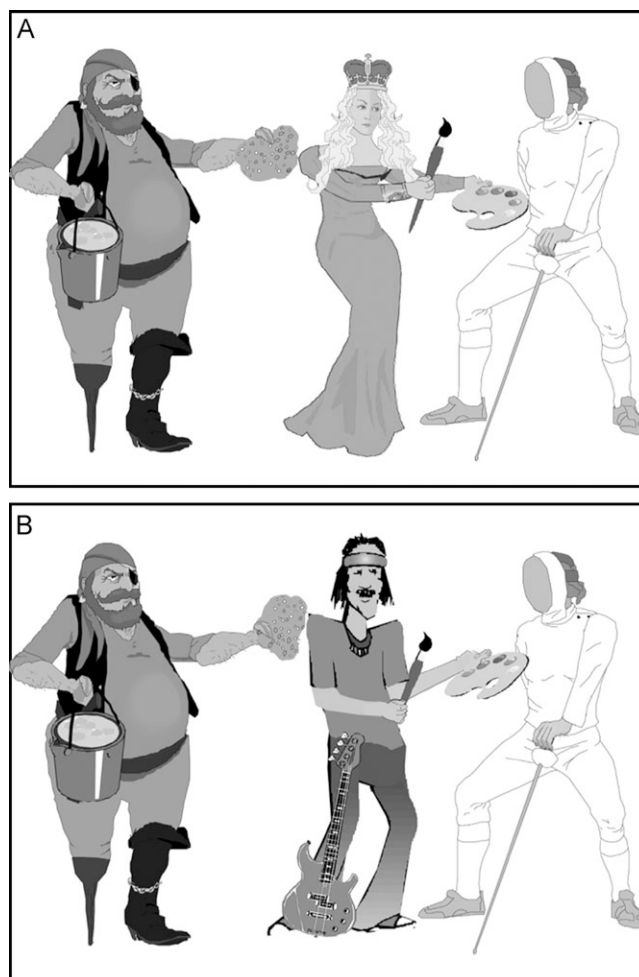


Figure 1. Images for an example item.

For masculine noun phrases in German, the grammatical function (subject vs. object) of the noun phrase is marked through nominative (*der*) and accusative case marking (*den*) on the determiner of the noun phrase. As a result, the masculine noun phrase was unambiguously marked as the subject (*der Musiker* “the musician”) or object (*den Musiker*, “the musician”) of the sentence. The scene for the unambiguous conditions (Fig. 1B) was created by replacing the princess in Figure 1A with a male character (musician). Crossing ambiguity (ambiguous and unambiguous) with canonicity (canonical and non-canonical) created 4 conditions.

For the initially ambiguous sentences (1a/b), the determiner of the sentence-final noun phrase marked that noun phrase as either the subject or the object of the sentence, resolving the local structural ambiguity (Experiments 1 and 2). Concurrently presented depicted events, however, provided role relations of who-does-what-to-whom for potentially earlier, verb-mediated disambiguation (Experiment 1). The verb in canonical sentences (*malt, paints* (1a)) identified the ambiguous character (the princess, Fig. 1A) as the agent of the depicted princess-painting-fencer event, whereas the verb in noncanonical sentences (*wäscht, washes* (1b)) identified the princess as patient of the pirate-washing-princess event. In contrast with Experiment 1, no scenes were present in Experiment 2. Participants thus had to rely on linguistic cues for structuring the sentence and for resolving the temporary structural ambiguity in sentences (1a) and (1b).

In addition to sentences (1a/b) and their unambiguous counterparts as well as the 2 images in Figure 1, each item contained a 2 further 4 sentences and 2 images for counterbalancing reasons (see Knoeferle et al. 2005). The original and counterbalancing versions of the sentences and images only differed in the role of the characters, whereas the verb and corresponding depicted actions remained unchanged. The counterbalancing ensured that each noun phrase/target character (the pirate and the fencer) was once the agent and once the patient and thus contributed to both the canonical and noncanonical condition. To illustrate this, the counterbalancing versions of sentences (1a/b), for instance, were “Die Prinzessin” (amb.) “wäscht den Pirat” (object) (*the princess* [amb.] *washes apparently the pirate* [object]) for the canonical condition and “Die Prinzessin” (amb) “malt offensichtlich der Fechter” (subject) (*the princess* [amb.] *washes apparently the fencer* [subject]) in the noncanonical condition. The corresponding counterbalancing image showed the princess washing the pirate while the fencer was depicted as painting the princess.

In the study by Knoeferle et al. (2005), there were 24 experimental items. Of these 24 items, we selected 20 and created another 20 images and sentences for the unambiguous condition as described above. To further increase the number of items for the ERP study, the character that was the referent of the first noun phrase in each of these 20 items (e.g., the princess for Fig. 1A and sentences (1a) and (1b)) was inserted into 3 of the other 20 items. In this way, an additional 3 items were created for each of the 20 items that we selected from the material set by Knoeferle et al. (2005). We used this way of generating additional stimuli because finding sufficient verbs that could be clearly depicted as actions proved difficult. Using only the depicted actions that Knoeferle et al. (2005) employed furthermore minimized the possibility that a potential absence of disambiguation effects might result from a change in materials.

There were 80 experimental items for the ERP study. Each participant saw an individually randomized list that contained 160 experimental and 200 filler trials. Each image was thus presented twice in a list, once in the original (Fig. 1A,B) and once in a counterbalancing version of the figures. Repetitions of an image and its counterbalanced version as well as repetitions that resulted from the above-described creation of the additional items were separated by at least 10 intervening trials. Experimental trials were separated by at least one, and experimental trials of the same condition were separated by at least 3 intervening trials.

Procedure

Prior to each trial, participants fixated a centrally located dot on the screen. Images were presented on a 19-inch monitor at a viewing distance of 80 cm. One second after image presentation onset, utterances were presented via speakers in 3 chunks (e.g., *The princess, paints apparently*, and *the fencer* (1a)). The 3 chunks derived from

recordings of complete utterances. The interstimulus interval between the chunks was 600 ms, giving a quasinatURAL flow of the sentence. We chose this manner of presentation for 2 reasons: First, it helped to avoid potential overlaps in the ERP components triggered at the verb and components that resulted from disambiguation through case marking on the determiner of the second noun phrase. Second, the ERPs to the different chunks show a fixed temporal relation to the accompanying picture stimulus. Each scene-sentence pair was followed by a pause that varied between 500 and 1100 ms. Participants were asked to sit still and minimize their eye movements during image presentation. They were instructed to attentively listen to the sentences. After half of the trials, participants had a short break. On 45 trials, participants answered a yes/no question about the presence or absence of an object in the scene. The questions were presented after a trial, always referred to the immediately preceding trial, and were randomly distributed over critical items and fillers. They ensured that people performed a comprehension task. Auditory stimuli of Experiment 2 were identical to Experiment 1, but no scene was presented. All other procedural details were identical to Experiment 1.

Recording and Analysis

The electroencephalogram (EEG) was recorded from 30 positions of the international 10-20 system including all 19 standard positions. An additional electrode was placed at the left mastoid as a reference. Vertical eye movements were measured with bipolar montages from an electrode above the left eyebrow and an electrode placed below the left orbital ridge. Two electrodes placed at the left and right external canthus measured horizontal eye movements. EEG data were recorded continuously using a bandpass of 0.01-70 Hz with a sampling rate of 250 Hz. The EEG was averaged for each experimental trial for epochs of 1024 ms including a 100-ms prestimulus baseline. Trials contaminated by eye movement artifacts were rejected off-line using individualized amplitude criteria determined by inspection of eye-blink artifacts. No more than 25 percent of the trials in each particular condition of a given participant were rejected because of eye-movement artifacts. We conducted analyses of variance (ANOVAs) on the mean amplitude of the average ERPs for the verb region (“verb,” e.g., *malt offensichtlich*) and the second noun phrase (“NP2,” e.g., *den Fechter*). The verb and noun phrase onsets were separate time-locking events for ERP averaging. On the basis of visual inspection and in correspondence with published results on the P600 component, the following time windows were chosen for the statistical analysis of the grand average ERPs: 500-800 ms relative to the onset of the verb and 500-800 ms relative to the onset of the second noun phrase. We first performed omnibus repeated measures ANOVAs with canonicity (canonical vs. noncanonical), ambiguity (ambiguous vs. unambiguous), anteriority (3 or 5 levels), and hemisphere (left vs. right electrodes) as factors. Findings of complex interactions were followed up by analyses comparing canonical and noncanonical conditions for ambiguous and unambiguous sentences separately. Separate sets of ANOVAs were conducted for midline (Fz, Cz, and Pz), parasagittal (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, and O2), and temporal (F7, F8, T7, T8, P7, and P8) electrode sets. For the midline analyses, the factor hemisphere was omitted. Huynh-Feldt adjustments to degrees of freedom were applied to correct for violation of the assumption of sphericity. We report the original degrees of freedom in conjunction with the corrected *P* values.

Results

Experiment 1 (Audiovisual)

Accuracy for the question-answering task was high (92.1%), suggesting participants understood sentences and images. Figures 2 and 3 show the grand average ERPs in the ambiguous and unambiguous conditions for frontal, central, and parietal midline electrodes from the onset of the verb and second noun phrase, respectively. The results of the corresponding statistical analyses are reported in Table 1.

Figure 2 illustrates a parietal positivity (P600) for initially ambiguous noncanonical versus canonical sentences at the verb

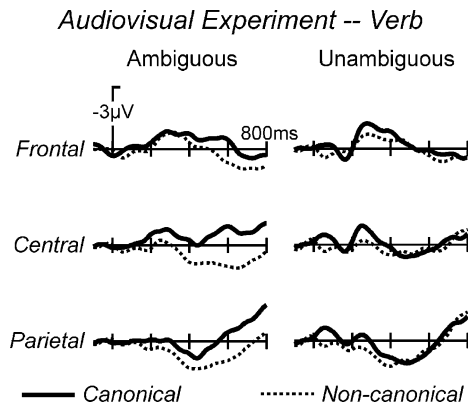


Figure 2. Grand average ERPs from the audiovisual experiment (Experiment 1) for 3 midline electrodes at the verb position. Waveforms were subjected to a digital lowpass (half amplitude cutoff 8 Hz) for visualization. A clear positivity emerges for ambiguous noncanonical relative to canonical sentences at the verb when disambiguating visual information is given.

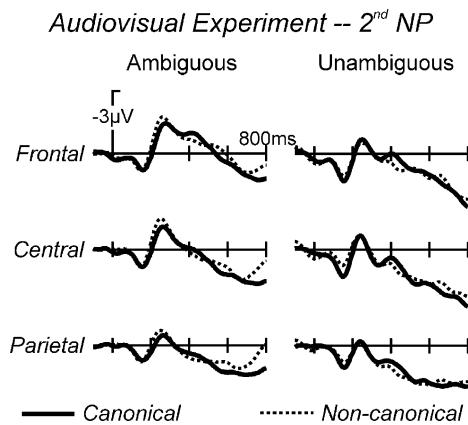


Figure 3. Grand average ERPs from the audiovisual experiment (Experiment 1) for 3 midline electrodes at the NP2 position.

Word position		Overall ANOVA			Ambiguous	Unambiguous
		Canonicity	Ambiguity	C × A	C	C
Verb	Midline	3.63##	0.01	9.23**	8.93**	0.02
	Parasagittal	1.77	0.15	4.91*	4.82*	0.03
	Temporal	0.93	0.39	0.11	1.01	0.41
NP2	Midline	3.49	6.07*	0.65	2.59	1.24
	Parasagittal	5.65*	12.89†	0.16	4.50##	2.83
	Temporal	7.44*	8.05*	0.01	4.36##	3.87##

Note: Given are the *F* values and degrees of freedom in each case *df* (1,15); main effects and interactions of factors hemisphere and anteriority are omitted for the sake of brevity; columns 2–4 show the results of the overall ANOVA; columns 5 and 6 show the main effect of canonicity for ambiguous and unambiguous sentences, respectively.

##*p* < 0.1.

**p* < 0.05.

***p* < 0.01.

†*p* < 0.005.

that identified relevant scene events. It further illustrates that there was no such positivity for noncanonical relative to canonical sentences during the corresponding time window in the unambiguous conditions (see Table 1). The fact that

a P600 for noncanonical relative to canonical sentences was observed in the ambiguous but not in the unambiguous conditions gave rise to canonicity by ambiguity interactions in the overall analysis. A main effect of canonicity was observed in the separate analyses for the ambiguous sentences.

The ERP pattern at the position of the second noun phrase is illustrated in Figure 3. At the position of the second noun phrase, crucially, no P600-like positivity for noncanonical relative to canonical sentences is observed in either ambiguous or unambiguous conditions (Fig. 3 and Table 1). For unambiguous sentences, this was expected because no ambiguity had to be resolved at this point. For ambiguous sentences, the absence of a P600-like component at the position of the second noun phrase corroborates the view that disambiguation had occurred earlier at the verb that mediated relevant depicted events.

Ambiguous sentences further show a more negative waveform, in particular, over frontal areas. The more negative amplitude in the ERP for ambiguous relative to unambiguous conditions gives rise to a main effect of ambiguity. The overall ANOVA furthermore confirmed a reliable effect of canonicity at parasagittal and temporal electrode sites. The canonicity effect resulted from a more negative waveform for noncanonical compared with canonical conditions. Previous ERP studies have also observed a more negative-going waveform for noncanonical relative to canonical sentences time locked to the onset of the second noun phrase at similar sites (see, e.g., Matzke et al. 2002, maximum at F7). Matzke et al. (2002) interpreted this negative deflection for the noncanonical relative to canonical condition as reflecting storage and retrieval procedures required for building the noncanonical structure. Based on their findings, we think that the main effect of canonicity in the overall analyses for the second noun phrase in our study (Table 1) may also reflect storage and retrieval procedures in building the non-canonical structure. This suggests that whereas depicted events trigger immediate ambiguity resolution at the verb, the revision process initiated by the depicted events did not entirely eliminate the increased demands of structural revision toward a noncanonical word order. In attempting to further examine whether the main effect of canonicity in the overall analyses resulted from a clear canonicity effect in both ambiguous and unambiguous conditions, we analyzed the data for these 2 conditions separately: crucially, the effect of canonicity was not significant in the separate analyses for ambiguous and unambiguous sentences, suggesting the canonicity effect is relatively weak.

Experiment 2

Accuracy on the test questions was high (91.2%). Figures 4 and 5 present the grand average ERPs in the ambiguous and unambiguous conditions for frontal, central, and parietal midline electrodes from the onset of the verb and second noun phrase, respectively. The results of the corresponding statistical analyses are reported in Table 2.

Figure 4 shows the absence of a P600 for ambiguous and unambiguous sentences at the verb position. The absence of a P600 for noncanonical compared with canonical sentences at the verb position in both ambiguous and unambiguous conditions is confirmed by the statistical analyses (Table 2).

In contrast, the ERPs elicited at the second noun phrase (Fig. 5) are associated with a P600 for the noncanonical relative to canonical sentences in the ambiguous conditions and, to a lesser extent, also for the unambiguous conditions. This pattern is

Auditory Experiment -- Verb

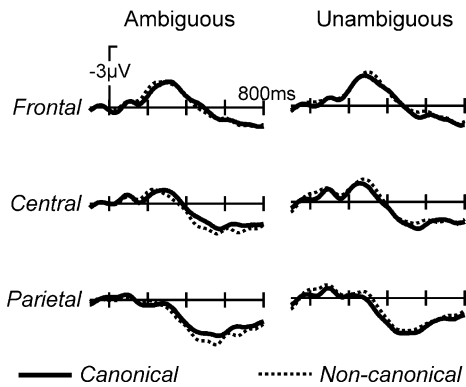


Figure 4. Grand average ERPs from the auditory control experiment for 3 midline electrodes at the verb position. No P600 activity is observed at this position for the noncanonical ambiguous sentences, as information for the disambiguation is not yet available at this position.

Auditory Experiment -- 2nd NP

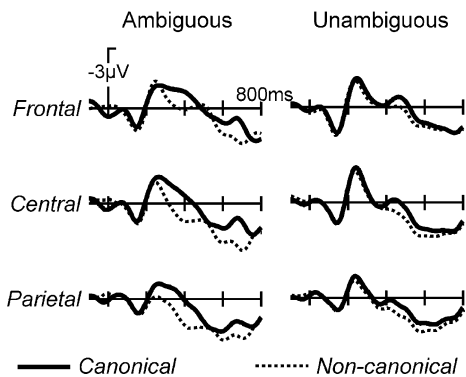


Figure 5. Grand average ERPs from the auditory control experiment (Experiment 2) for 3 midline electrodes at the NP2 position. A clear P600 emerges for the noncanonical ambiguous sentences, as case marking information identifying the noun phrase as the subject of the sentence becomes available at this position.

confirmed by the statistical analyses for the NP2 position (Table 2). Observing a P600 at the second noun phrase for unambiguous sentences in Experiment 2 suggests that people sometimes disregarded the determiner on the first noun phrase that marked that noun phrase as the object. The absence, however, of a P600 on the second noun phrase for noncanonical relative to canonical sentences in the unambiguous conditions of Experiment 1 together with the P600 for noncanonical versus canonical sentences that we observed for that region in the ambiguous conditions of Experiment 2 provide a valid baseline for interpreting the P600 at the verb for the ambiguous sentences in Experiment 1.

General Discussion

The auditory ERP recordings show that visual scenes trigger immediate structural revision of locally structurally ambiguous utterances. Evidence for this claim comes from a P600 for ambiguous noncanonical versus canonical sentences in Experiment 1, time locked to the verb that made available the relevant depicted events. During the same early time window (the verb), we neither found a P600 for structurally unambiguous controls

Table 2

Auditory experiment, verb and NP2 positions, statistical results

Word position		Overall ANOVA			Ambiguous	Unambiguous
		Canonicity	Ambiguity	C × A	C	C
Verb	Midline	0.13	2.38	2.56	1.91	0.59
	Parasagittal	0.01	1.85	3.18##	1.58	1.37
	Temporal	0.11	3.02	3.06	1.63	2.38
NP2	Midline	16.84***	0.45	1.01	8.54*	6.41*
	Parasagittal	6.54*	0.97	0.59	6.53*	3.16##
	Temporal	8.18*	0.71	0.91	5.21*	3.97##

Note: Given are the *F* values and degrees of freedom in each case *df* (1,15); main effects and interactions of factors hemisphere and anteriority are omitted for the sake of brevity; columns 2–4 show the results of the overall ANOVA; columns 5 and 6 show the main effect of canonicity for ambiguous and unambiguous sentences, respectively.

p < 0.1.

* *p* < 0.05.

** *p* < 0.01

♯ *p* < 0.005.

*** *p* < 0.001.

in scenes (Experiment 1) nor for ambiguous or unambiguous utterances in the absence of scenes (Experiment 2). Rather, the data from Experiment 2 show that when scenes are absent, disambiguation occurs through linguistic cues (case marking) on the determiner of the sentence-final noun phrase for structurally ambiguous noncanonical relative to canonical sentences (P600).

The data from these 2 experiments have important methodological and theoretical implications. From a methodological viewpoint, the electrophysiological data from Experiment 1 support the interpretation of gaze patterns in the studies by Knoeferle et al. (2005) as reflecting rapid structural revision. Furthermore, correlating these 2 different measures for related underlying comprehension processes informs us about the neuronal processes underlying the use of scene information, the semantic interpretation that listeners pursue, and the kinds of information in scenes that they attend to and exploit. Although electrophysiological methods alone may in the future reveal both attention in scenes and the neuronal processes underlying comprehension (e.g., Joyce et al. 2002), this has not yet been investigated for spoken sentence comprehension in relatively complex scenes.

From a theoretical and neurocognitive perspective, the findings from Experiments 1 and 2 crucially extend existing insights into the neural correlates underlying the disambiguation of local structural ambiguity. First, our findings for auditory presentation in the absence of scenes (Experiment 2) showed that the effects that Matzke et al. (2002) observed during the disambiguation of locally ambiguous German subject-verb-object/object-verb-subject sentences in reading generalize to the comprehension of spoken sentences (see also Osterhout and Holcomb 1993).

Second, findings from the audiovisual experiment (Experiment 1) provide important insights into the role of scene information for incremental structural disambiguation: nonlinguistic (depicted events)—just as linguistic (case marking)—cues trigger rapid syntactic reanalysis. Our findings of verb-mediated disambiguation in the audiovisual experiment are clearly compatible with interactionist models of sentence comprehension (e.g., Altmann and Steedman 1988; MacDonald et al. 1994; Trueswell and Tanenhaus 1994) and emphasize that scene information should explicitly be included into these

models to account in more detail for the nature of its influence on processes of structural revision. Knoeferle and Crocker (forthcoming), for instance, propose a detailed processing account of the temporal interplay between scene information and utterance comprehension mechanisms.

With respect to the role of scene information in disambiguation, it is further interesting to note that the distribution of the P600-like component was similar—whether disambiguation was triggered by depicted events at the verb as in the ambiguous conditions of Experiment 1 or by linguistic marking on the second noun phrase for the ambiguous conditions of Experiment 2 (Fig. 6). This suggests that the kinds of cues (e.g., nonlinguistic depicted events vs. linguistic cues such as case marking) that trigger disambiguation do not fundamentally modulate the neural correlates underlying disambiguation mechanisms.

This view is further corroborated by existing findings on structural disambiguation through linguistic cues. Beim Graben et al. (2000) examined the comprehension of German wh-questions in which a noun following the wh-word was ambiguous between a subject (canonical) and an object (non-canonical). Disambiguation of the initial ambiguity, just as in Experiment 1 of the present paper, took place at the verb. However, unlike Experiment 1, disambiguation was triggered by number agreement between the initial noun phrase and the verb rather than by depicted events. Beim Graben et al. (2000) observed a P600 at midline electrodes in response to disambiguation through number agreement. Furthermore, Frisch et al. (2002) reported a P600 at the midline when case marking on the determiner of a noun phrase disambiguated a local subject-object/object-subject ambiguity. Findings from these existing studies in which disambiguation occurred through linguistic cues (beim Graben et al. 2000; Frisch et al. 2002; Matzke et al. 2002) together with the results from our experiments provide strong evidence for the view that the kinds of cues—linguistic (case marking and number agreement) versus nonlinguistic (depicted events)—triggering disambiguation do not modulate the neural correlates of structural disambiguation in a fundamental manner.

The fact that we observed a positivity with a peak latency of 600 ms and similar topography for utterance- and scene-based disambiguation strongly suggests that the P600 is sensitive to the integration of a variety of information sources. This proposal is compatible with existing neurocognitive models (e.g., Friederici 2002; Hagoort 2003) insofar as those view the P600 as a late component the generation of which may involve the

integration of syntactic, semantic, and pragmatic information. What our findings add to these accounts is the insight that scene information should be included among the informational sources that participate in the revision and integration processes reflected by the P600.

Indeed, disambiguation through depicted events presumably requires cross-modal integration of visual representations (e.g., from event scenes) with representations from the unfolding utterance. The fact that we observed a P600 with similar latency and topography in response to both scene-based (cross-modal) and utterance-based (unimodal) disambiguation fits well with existing proposals that together suggest a tentative link between the P600, the posterior superior temporal gyrus, and cross-modal information processing (see, e.g., Friederici et al. 2003; Hagoort 2003; Indefrey 2004). Friederici et al. (2003) draw a tentative link between semantic and syntactic integration processes and increased activation in the posterior portion of the left superior temporal gyrus (encompassing Brodmann's areas 22, 39, and 40). Interestingly, the posterior superior temporal gyrus has also been identified as the locus for cross-modal (e.g., audiovisual) information integration (see, e.g., Wright et al. 2003; Spitsyna et al. 2006; see also Hagoort 2003). Together these findings further emphasize the view that the P600 component is truly domain general (e.g., Coulson et al. 1998; Münte et al. 1998; Patel et al. 1998).

A further interesting question regarding the neurocognitive ramifications of our findings is whether the P600-like component observed at the verb in the audiovisual experiment is in fact a P300 component and specifically a P3b. The P3b is a positivity with a maximum over centroparietal sites and belongs to the P300 family—a host of domain-general components. The P3b is, for instance, observed in the oddball paradigm, where people listen to a sequence of frequent tones interspersed with infrequent deviant tones, and is elicited in response to the rare deviant stimulus. It has been described as reflecting the resolution of uncertainty and the surprise associated with a given task-relevant stimulus (e.g., Kutas 1977; Picton 1992; see Coulson et al. 1998).

In light of the design, materials, and the task for Experiment 1, a P3b interpretation of the P600-like component that we observed appears plausible. People may have cued into the fact that relating the utterances to the scene events facilitates the comprehension task. When disambiguation toward a non-canonical structure occurred, the rarity of either a word in the utterance (e.g., a subject case-marked determiner on the sentence-final noun phrase) or of noncanonical event relations may have elicited a P3b. This is plausible because noncanonical object-initial sentences in our study were less frequent than canonical sentences. The P3b would in this case reflect surprise associated with the discovery that—based on either linguistic cues or depicted events—a canonical subject-verb-object structure and corresponding agent-action-patient representations cannot be built. A P3b interpretation of scene-based structural disambiguation in particular is plausible because the P3b is a domain-general component and thus likely sensitive to the processing of scene information during language processing.

Indeed, a key question is whether the P600 should not best be described as a member of the P300 family (see, e.g., Coulson et al. 1998), a view that is supported by the similar scalp distributions of these 2 components. The present findings alone cannot decide this issue. Based on the observation, however, that the utterance-based and scene-based positivities are highly

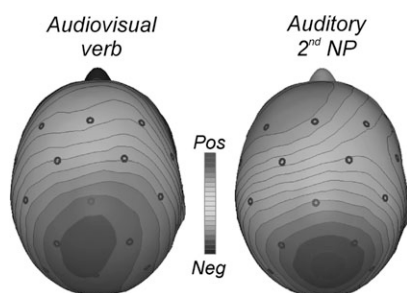


Figure 6. Spline interpolated isovoltage maps depicting the canonical minus noncanonical difference at 400 ms after stimulus onset for the ambiguous sentences. The positivity at the verb position in the audiovisual experiment and its counterpart at the second NP position in the auditory experiment have a virtually identical scalp distribution.

similar in both peak latency and topography, it appears that if one of these 2 observed components is a member of the P300 family, the other one likely also belongs to this host of positive domain-general components. In all likelihood, our findings thus corroborate existing proposals that describe the positivities observed for structural violations as a domain-general response (see, e.g., Münte et al. 1998; Patel et al. 1998; but see Osterhout et al. 1996).

Regardless of whether the component that we observed for scene- and utterance-based disambiguation toward a noncanonical structure is a P600 or a more general P3, our findings support equality between scene-based versus linguistic cues in incremental structural disambiguation. Clearly, scene context can be exploited on par with a linguistic context for the incremental structuring of a sentence.

Funding

German research foundation Deutsche Forschungsgemeinschaft (PhD scholarship GRK-715 to PK, postdoctoral fellowship to PK, SFB-378-“ALPHA” to MWC, and MU1311/13-1 to TFM).

Notes

Conflict of Interest: None declared.

Address correspondence to Dr Pia Knoeferle, Center for Research in Language, 9500 Gilman Drive, Department 0526, Room CSB 263, La Jolla, CA, 92093-0526, USA. Email: pknoeferle@ucsd.edu.

References

- Altmann GTM, Kamide Y. 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*. 73: 247–264.
- Altmann GTM, Steedman M. 1988. Interaction with context during human sentence processing. *Cognition*. 30:191–238.
- beim Graben P, Saddy JD, Schlesewsky M. 2000. Symbolic dynamics of event-related brain potentials. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*. 62:5518–5541.
- Chambers CG, Tanenhaus MK, Magnuson JS. 2004. Actions and affordances in syntactic ambiguity resolution. *J Exp Psychol Learn Mem Cogn*. 30:687–696.
- Coulson S, Kutas M, King JW. 1998. Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang Cogn Process*. 13:21–58.
- Friederici AD. 2002. Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*. 6:78–84.
- Friederici AD, Pfeifer E, Hahne A. 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Brain Res Cogn Brain Res*. 1:183–192.
- Friederici AD, Rüschemeyer SA, Hahne A, Fiebach CJ. 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex*. 13:170–177.
- Frisch S, Schlesewsky M, Saddy D, Alpermann A. 2002. The P600 as an indicator of syntactic ambiguity. *Cognition*. 85:B83–B92.
- Hagoort P. 2003. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*. 20:S18–S29.
- Hagoort P, Brown CM, Groothusen J. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang Cogn Process*. 8:439–483.
- Indefrey P. 2004. Hirnaktivierung bei syntaktischer Sprachverarbeitung: eine Meta-Analyse. In: Müller HM, Rickheit G, editors. *Neurokognition der Sprache*. Tübingen (Germany): Stauffenberg Verlag. p. 31–50.
- Joyce CA, Gorodnitsky I, King JW, Kutas M. 2002. Tracking eye fixations with electroocular and electroencephalographic recordings. *Psychophysiol*. 39:607–618.
- Knoeferle P, Crocker MW. 2006. The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cogn Sci*. 30:481–529.
- Knoeferle P, Crocker MW. Forthcoming. The influence of recent scene events on spoken comprehension: evidence from eye tracking. *J Mem Lang*.
- Knoeferle P, Crocker MW, Scheepers C, Pickering MJ. 2005. The influence of the immediate visual context on incremental thematic role assignment: evidence from eye movements in depicted events. *Cognition*. 95:95–127.
- Kutas M, McCarthy G, Donchin E. 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*. 197:792–795.
- MacDonald MC, Pearlmutter NJ, Seidenberg MS. 1994. The lexical nature of syntactic ambiguity resolution. *Psychol Rev*. 101:676–703.
- Matzke M, Mai H, Nager W, Rüsseler J, Münte T. 2002. The costs of freedom: an ERP study of non-canonical sentences. *Clin Neurophysiol*. 113:844–852.
- Münte T, Heinze H, Matzke M, Wieringa BM, Johannes S. 1998. Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia*. 39:66–72.
- Osterhout L, Holcomb P. 1993. Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Lang Cogn Process*. 8:413–488.
- Osterhout L, Holcomb PJ, Swinney DA. 1994. Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J Exp Psychol Learn Mem Cogn*. 20:786–803.
- Osterhout L, McKinnon R, Bersick M, Corey V. 1996. On the language specificity of the brain response to syntactic anomalies: is the syntactic positive shift a member of the P300 family? *J Cogn Neurosci*. 8:507–526.
- Patel AD, Gibson E, Ratner J, Besson M, Holcomb P. 1998. Processing syntactic relations in language and music: an event-related potential study. *J Cogn Neurosci*. 10:717–733.
- Picton TW. 1992. The P300 wave of the human event-related potential. *J Clin Neurophysiol*. 9:456–479.
- Schlesewsky M, Fanselow G, Kliegl R, Krems J. 2000. The subject preference in the processing of locally ambiguous wh-questions in German. In: Hemforth B, Konieczny L, editors. *German sentence processing*. Dordrecht (Germany): Kluwer. p. 65–94.
- Sedivy JC, Tanenhaus MK, Chambers CG, Carlson GN. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition*. 71:109–148.
- Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS. 2006. Converging language streams in the human temporal lobe. *J Neurosci*. 26:7328–7336.
- Spivey MJ, Tyler MJ, Eberhard KM, Tanenhaus MK. 2001. Linguistically mediated visual search. *Psychol Sci*. 12:282–286.
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard K, Sedivy JC. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*. 268:632–634.
- Trueswell JC, Tanenhaus MK. 1994. Towards a lexicalist framework for constraint-based syntactic ambiguity resolution. In: Clifton C, Frazier L, Rayner K, editors. *Perspectives in sentence processing*. Hillsdale (NJ): Lawrence Erlbaum Associates. p. 155–179.
- Wright TM, Pelphrey KA, Truett A, McKeown MJ, McCarthy G. 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex*. 13:1034–1043.