The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension

CrossMark

Neil Cohn [a,c,*], Ray Jackendoff [b,a], Phillip J. Holcomb [a], Gina R. Kuperberg [a,d]

[a] Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, USA
[b] Department of Philosophy and Center for Cognitive Studies, Tufts University, 115 Miner Hall, Medford, MA 02155, USA
[c] Department of Cognitive Science, University of California, San Diego, 9500 Gilman Dr. Dept 0526, La Jolla, CA 92093-0526, USA
[d] Department of Psychiatry and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Bldg 149, 13th Street, Charlestown, MA 02129, USA

ARTICLE INFO

ABSTRACT

Constituent structure has long been established as a central feature of human language. Analogous to how syntax organizes words in sentences, a narrative grammar organizes sequential images into hierarchic constituents. Here we show that the brain draws upon this constituent structure to comprehend wordless visual narratives. We recorded neural responses as participants viewed sequences of visual images (comics strips) in which blank images either disrupted individual narrative constituents or fell at natural constituent boundaries. A disruption of either the first or the second narrative constituent produced a left-lateralized anterior negativity effect between 500 and 700 ms. Disruption of the second constituent also elicited a posteriorly-distributed positivity (P600) effect. These neural responses are similar to those associated with structural violations in language and music. These findings provide evidence that comprehenders use a narrative structure to comprehend visual sequences and that the brain engages similar neurocognitive mechanisms to build structure across multiple domains.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Constituent structure is a hallmark of human language. Discrete units (words) group into larger constituents (phrases), which can recursively combine in indefinitely many ways (Chomsky, 1965; Culicover & Jackendoff, 2005). Language, however, is not our only means of communication. For millennia, humans have told stories using sequential images, whether on cave walls or paintings, or in contemporary society, in comics or films (Kunzle, 1973; McCloud, 1993). Analogous to the way words combine in language, individual images can combine to form larger constituents that enable the production and comprehension of complex coherent visual narratives (Carroll & Bever, 1976; Cohn, 2013b; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; Gernsbacher, 1985; Zacks, Speer, & Reynolds, 2009).

It has long been recognized that narratives follow a particular structure (Freytag, 1894; Mandler & Johnson, 1977). This dates all the way back to Aristotle's observations about plot structure in theater (Butcher, 1902). We have recently formalized a narrative grammar of sequential images, in which each image plays a

categorical role based on its narrative function within the overall visual sequence (Cohn, 2013b, 2014). These image units can subsequently group together to form *narrative constituents*, which themselves fulfill narrative roles in the overall structure. While this general approach is similar to previous grammars of discourse and stories (e.g., Clark, 1996; Hinds, 1976; Labov & Waletzky, 1967; Mandler & Johnson, 1977; Rumelhart, 1975), it differs from these precedents in the simplicity of its recursive structures (Cohn, 2013b), the incorporation of modifiers beyond a canonical narrative arc (Cohn, 2013a, 2013b), and the explicit separation of structure and meaning (Cohn et al., 2012), see Cohn (2013b) for more details.

To better understand this narrative grammar, consider the sequence shown in Fig. 1. This has two narrative constituents. The first constituent contains two images: the first image plays the narrative role of an "Initial," functioning to set up the central event ("hitting the ball"), while the second image plays the narrative role of a "Peak" as it depicts the hitting action itself. The second constituent consists of four images: the first, an "Establisher," functions to introduce the characters involved in the main event; the second, an "Initial," sets up the event; the third functions as a "Peak" depicting the climactic crashing event itself, and the fourth image acts as a "Release," resolving this central action. Importantly, these two narrative constituents are related on a higher level of narrative structure, such that the first larger constituent

* Corresponding author at: Department of Cognitive Science, University of California, San Diego, 9500 Gilman Dr. Dept. 0526, La Jolla, CA 92093-0526, USA. Tel.: +1 858 822 0736; fax: +1 858 822 5097.

E-mail address: neilcohn@visuallanguagelab.com (N. Cohn).
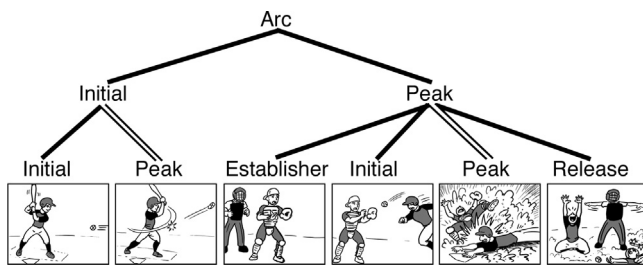
Fig. 1. Narrative structure of a visual sequence. This sequence contains two narrative constituents. The first two panels together depict an event; the first panel is an "Initial," which sets up the climatic event in the second panel, a "Peak." In turn, these two panels together serve as an Initial for the sequence as a whole, and the event depicted by the final four panels serves as Peak for the entire sequence.

functions as an "Initial" to set up the content of the second larger constituent, which itself acts as the climactic "Peak" of the whole sequence. In more complex narratives, embedding along similar lines can be deeper or altered through modifiers.

In previous experimental research, we have shown that these narrative categories follow distributional trends in sequences, relying on cues from both image content as well as their context within a sequence (Cohn, 2014). In addition, our previous work suggests that, during the comprehension of visual narrative sequences, the brain uses this narrative structure in combination with more general semantic schemas (Schank & Abelson, 1977) to build up global narrative coherence, which, in turn, facilitates semantic processing of incoming panels (Cohn et al., 2012). So far, however, it remains unclear how the brain responds to input that actually violates expectations that are based on our representation of this narrative structure. Addressing this question was the aim of the present study. We show that the constituent structure in our proposed narrative grammar is not just an interesting theoretical construct: it can be detected experimentally.

The paradigm we developed is modeled on classic psycholinguistic experiments that demonstrated that word-by-word comprehension engages grammatical constituent structure. In an important series of behavioral studies, participants listened to simple sentences such as *My roommate watched the television*, during which there was a burst of white noise (a "click": depicted here as **). Initial research using this paradigm showed that clicks appearing *within* a syntactic constituent (e.g., disrupting the noun-phrase: *My ** roommate watched…*) were recalled less accurately than clicks appearing *between* syntactic constituents (e.g., between the noun-phrase and the verb-phrase: *My roommate ** watched…*), and that false recollection of clicks remembered them as occurring between constituents (Fodor & Bever, 1965; Garrett & Bever, 1974). Later studies using online monitoring tasks found that reaction times were faster to clicks placed between constituents than those within syntactic constituents, and faster to those within first constituents than second constituents (Abrams & Bever, 1969; Bond, 1972; Ford & Holmes, 1978). The success of this "structural disruption" technique as a method of examining grammatical structure in language has led to its use beyond the study of structure in language, to study structure in music (Berent & Perfetti, 1993; Kung, Tzeng, Hung, & Wu, 2011) and visual events (Baird & Baldwin, 2001).

At a neural level, studies using event-related potentials (ERPs) have reported two effects in association with structural (syntactic) aspects of language processing: (1) a left-lateralized anterior negativity (Friederici, 2002; Friederici, Pfeifer, & Hahne, 1993; Hagoort, 2003; Neville, Nicol, Barss, Forster, & Garrett, 1991), starting at or before 350 ms, and (2) a posteriorly-distributed positivity (P600), starting at around 500 ms, although sometimes earlier (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). The left anterior

negativity effect tends to be evoked by words that are consistent (versus inconsistent) with one of just two or three possible upcoming syntactic structures predicted by the context (Lau, Stroud, Plesch, & Phillips, 2006), and it is seen even when this context is semantically non-constraining (Gunter, Friederici, & Schriefers, 2000) or semantically incoherent (Münte, Matzke, & Johannes, 1997). The P600 is most likely to be triggered when an input violates a strong, high certainty single structural expectation established by a context, particularly when this context is also semantically constraining (Kuperberg, 2007, 2013). It is believed to reflect prolonged attempts to make sense of the input (Kuperberg, 2007, 2013; Sitnikova, Holcomb, & Kuperberg, 2008a). Notably, both the left anterior negativity and P600 effects are distinct from the well-known N400 effect—a widespread negativity between 300 and 500 ms that is modulated by both words (Kutas & Hillyard, 1980) and images (Barrett & Rugg, 1990; Barrett, Rugg, & Perrett, 1988) that match versus mismatch contextual expectations about the semantic features of upcoming input, rather than expectations about its grammatical structure (Kutas & Federmeier, 2011).

Here, we ask whether violations of narrative constituent structure in sequential images produce neural effects analogous to those seen in response to structural violations in language. We developed a structural disruption paradigm, analogous to the classic "click" paradigm that, as discussed, provided early evidence that comprehenders use a syntactic constituent structure to comprehend language (Fodor & Bever, 1965; Garrett & Bever, 1974). Our paradigm also shares similarities with so-called ERP "omission" paradigms in which, rather than examining the neural response to a stimulus that is incongruous (versus congruous) with a context, ERPs are time-locked to the *omission* of the expected stimulus. We have known since the late 1960s that the omission of expected stimuli can evoke a large brain response (Klinke, Fruhstorfer, & Finkenzeller, 1968; Simson, Vaughan, & Walter, 1976), and more recently, this phenomenon has been interpreted within a generative Bayesian predictive coding framework (Clark, 2013; Friston, 2005; Rao & Ballard, 1999). According to this framework, the brain constructs an internal model of the environment by constantly assessing incoming stimuli in relation to their preceding context and stored representations. Top-down predictions are compared, at multiple levels of representation, with incoming stimuli, and the difference in the neural response between the top-down prediction and the bottom-up input—the "prediction error"—is passed up to a higher level of representation, where it is used to adjust the internal model or, when the input violates a very high certainty expectation, switch to an alternative model that can better explain the combination of the context and the incoming stimulus. Neural responses to omissions are taken to reflect pure neural prediction error, produced by the mismatch between top-down predictions and (absent) bottom-up input (Bendixen, Schröger, & Winkler, 2009; Friston, 2005; Todorovic, van Ede, Maris, & de Lange, 2011; Wacongne et al., 2011).

In our paradigm, participants viewed six-panel-long wordless visual sequences, presented image-by-image. These panels were constructed to have two narrative constituents in various different structural patterns (see Section 2). In some of the visual sequences, we inserted "blank" white panels devoid of content ("omission" stimuli). The blank panels fell either *within* a narrative constituent (either the first or the second constituent) or *in between* the two narrative constituents (see Fig. 2 for an example). Importantly, because we used several patterns of constituent structure, with narrative boundaries located after panel 2, 3, or 4, blank panels could appear anywhere from the second to fifth panel position in the sequence. This meant that comprehenders could not use ordinal position as a direct cue to predict when a blank panel would occur. We measured ERPs to these blank panels, and compared the ERP response produced by those that fell *within*
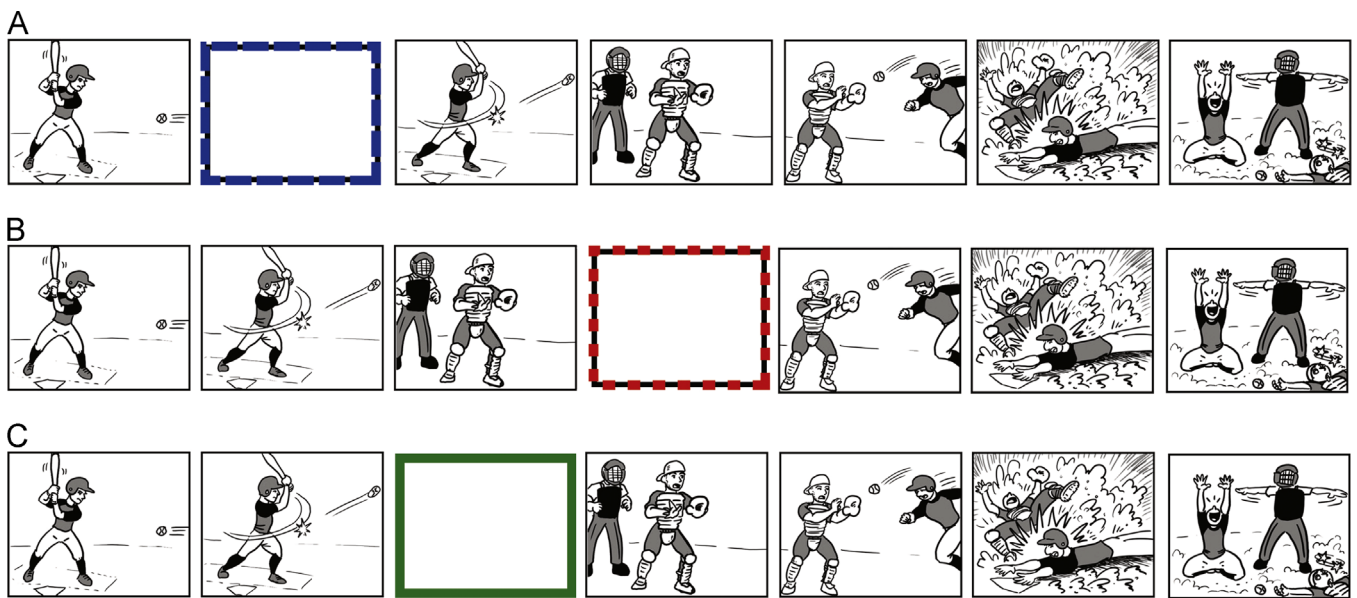
Fig. 2. Sample experimental stimuli. Experimental conditions were created by inserting white "blank images" either a) within the first (WC1), b) within the second (WC2) constituent, or c) between the two constituents at a natural constituent boundary (BC). This example sequence locates the constituent boundary after the second panel. However, we varied the position of the constituent boundary across stimuli, such that in others the boundary appeared after the third or fourth panel.

narrative constituents (disrupting a narrative constituent) to those produced by blank images that fell at the natural break *between* narrative constituents.

The logic of this design is as follows: if viewers use a narrative constituent grammar to guide sequential image comprehension, blank images that disrupt this narrative structure should produce a prediction error, as detected by an ERP response. This is in line with the reasoning behind ERP "omission" paradigms in which the brain activity to the omission of an expected stimulus provides evidence for its anticipation, as discussed above. Importantly, this disruption should be greatest when the blank image falls *within* a narrative constituent, directly disrupting the narrative structure, than when it falls in-between two constituents at a natural constituent boundary, leading to a larger prediction error and a larger ERP response to within-constituent blanks than between-constituent blanks. Such an effect would provide direct evidence that comprehenders use a stored narrative structure to anticipate upcoming aspects of structure. Moreover, if narrative structure engages neurocognitive mechanisms similar to those for language comprehension, these ERP responses might manifest as a left-lateralized anterior negativity and a P600 effect.

2. Methods

2.1. Stimuli

Novel 6-frame long comic strips were created (135) to have two narrative constituents. These sequences drew from a corpus of panels culled from six volumes of the *Complete Peanuts* by Charles Schulz (1952–1968). We used *Peanuts* comics because they have 1) consistent panels sizes; 2) characters and situations which are familiar to most people; and 3) a large corpus of strips to draw from. In order to eliminate any effects of written language, we only used panels without text, or digitally deleted the text from panels. Our creation of novel sequences ensured that they adhered to the constituent structures required for our experiment, and that participants would not be familiar with the specific sequences used in the experiment. Boundaries between constituents followed theoretically defined criteria (Cohn, 2013b), and were confirmed using a behavioral "segmentation task" (Gernsbacher, 1985) in which 20 participants (15 male, 5 female, mean age: 22) drew lines between images in each sequence that most intuitively divided them into two parts. Our final 120 strips had a 71% agreement for the location of the boundary between constituents, which appeared after image 2, 3, or 4 (40 of each

type) resulting in three different patterns of constituent structures throughout the stimuli.

For each sequence, we created three experimental conditions. We introduced blank images that disrupted either the first or the second narrative constituent (Within Constituent Blanks: WC1 and WC2 respectively), or alternatively, that occurred between the two narrative constituents (Between Constituents: BC), as depicted in Fig. 2. Because our sequences used three different patterns of constituent structures that varied the location of the narrative boundary, disruptions could therefore appear anywhere from the second to fifth positions across all sequences and all disruption types. This variation was introduced to avoid any systematic confound between ordinal position and position of the blank panels. By definition the WC2 blanks, on average, appeared later than the BC blanks which appeared later than the WC1 blanks: WC1 blanks (average position: 3; range: 2–4); BC blanks (average position: 4; range: 3–5), WC2 blanks (average position: 5; range: 4–6). However, because, across the stimulus set, a blank panel could appear at the same ordinal position in the WC1, WC2 and the BC conditions, participants could not use ordinal position as a cue during the experiment to predict when the blank would occur.

Our randomized lists also included filler sequences that had no blank images, two blank images in a row (to prevent normal images from always following a disruption), and/or a coherence violation (verified in a previous rating study). In total, each participant saw an equal number of strips with blanks (90 single-blanks+15 double-blanks=105) and without blanks (105 no-blank sequences), and 60 sequences with coherence violations and 150 fully coherent images.

2.2. Participants

Twenty-five right-handed, English speaking comic readers with normal vision (8 male, 17 female, mean age=19.9) were paid for participation and gave their informed written consent following the guidelines of the Tufts University Internal Review Board. Comic reading "expertise" was assessed using the "Visual Language Fluency Index" (VLFI) questionnaire which asks participants to assess the frequency with which they read and draw different types of visual narratives (comic books, comic strips, Japanese manga, and graphic novels), both currently and while growing up (see Cohn et al. (2012) for more details). Our prior work has shown that the "VLFI score" resulting from this metric correlates with both behavioral and neurophysiological effects in visual narrative comprehension (Cohn et al., 2012). An idealized average VLFI score would be 12, with low being below 7 and high above 20. In this study, participants' overall mean VLFI score was 16.12 (SD=6.7, range: 7.6–32).

2.3. Procedure

Participants viewed each sequence image-by-image on a computer screen while we measured their EEG. Each panel in the sequence depicted a black-and-white image on a white square that was centered on a black screen. The blank panels simply depicted a white square (of the same dimensions) also centered on

the black screen. Lights were kept on throughout the experiment to avoid a "flashing effect" that induced blinks. Images with pictorial content remained on the screen for 1350 ms, whereas blank panels appeared for 750 ms in order to appear as distinctly separate in character from narrative images (i.e., they were not panels with missing information). To reinforce this interpretation, participants were told in advance that sequences might contain blank frames, and that these images were not part of the narrative. An ISI of 300 ms of a full black screen separated each panel (both the pictorial panels and the blank panels) in order to prevent a "layering" effect of sequential images appearing like a flipbook animation. Participants rated the coherence of each sequence on a 1 to 5 scale for how easy it was to understand (1=difficult, 5=easy).

2.4. Data analysis

ERPs were time-locked to the onset of each blank image and collapsed across images. EEG trials with eye-blinks, eye-movements, artifact caused by muscle movements, and/or artifact caused by signal loss or blocking (i.e., a flat line) were rejected from analysis. Rejection of trials was determined by visual inspection of raw data for each participant, and rejection rates were kept below 15% for each stimulus event (i.e., maintaining at least 26 of 30 trials per condition per participant). We excluded one participant's data for exceeding this maximum rejection rate. The remaining trials maintained after artifact rejection were used in our averaged ERPs.

We analyzed mean voltages within the epochs of 300–500 ms and 500–700 ms. The electrode montage used 29 electrodes organized into midline and peripheral regions of interest for analysis of ERP data (see Fig. 3). The midline regions consisted of prefrontal (FPz, FP1, FP2), frontal (Fz, FC1, FC2), central (Cz, C3, C4), parietal (CP1, CP2, Pz), and occipital (O1, O2, Oz) regions. Peripheral regions included the left frontal (F3, F7, FC5), right frontal (F4, F8, FC6), left posterior (CP5, T5, P3), and right posterior (CP6, T6, P4) regions. Electrodes below the left eye (LE) and next to the right eye (HE) recorded blinks and eye movements. All electrodes were referenced to an electrode placed on the left mastoid (A1), while differential activity was monitored in the right mastoid (A2). An SA Bioamplifier amplified the electroencephalogram (EEG) using a bandpass of .01–40 Hz and continuously sampled at a rate of 200 Hz. Electrode impedances were kept below 10 kΩ for the eyes and below 5 kΩ at all other sites.

3. Results

Participants rated the coherence of each whole sequence on a five-point scale (1=hard to understand, 5=easy to understand). Position of the blank image had no effect on participants' overall coherence ratings (BC=3.7 (.32); WC1=3.8 (.33); WC2=3.8 (.38); $F(2,48)$=.792, p=.459).

Our analysis of ERP effects to blank images found no main effects or interactions in the 300–500 ms epoch (a post-hoc analysis looking at the 300–400 ms and 400–500 ms separately also showed no effects in the 300–400 ms time window, although, as noted below, it did reveal some effects between 400 and 500 ms). However, in the 500–700 ms epoch, omnibus repeated-measures ANOVAs with a Bonferroni correction for multiple comparisons showed significant interactions between Position (BC, WC1, WC2) and Region across mid-regions of the scalp (see Fig. 4), $F(8,184)$=3.39, $p < .05$, as well as between Position, Region, and Hemisphere at peripheral regions, $F(2,46)$=5.72, $p < .01$. Follow up analyses of these interactions at individual regions revealed two distinct ERP effects, distinguished by differences in polarity and scalp distribution (see Fig. 4): a left-lateralized anterior negativity effect, and a posteriorly distributed positivity effect.

In the left frontal region, WC1 blank images evoked a larger negativity than BC blanks, $F(1,23)$=4.93, $p < .05$. WC2 blank images evoked a larger negativity than BC blanks in the same left frontal region, $F(1,23)$=5.25, $p < .05$, as well as at the prefrontal region, $F(1,23)$=8.37, $p < .01$. A post-hoc analysis between 400 and 500 ms in the left anterior region indicated a near-significant WC2-BC effect, $F(1,17)$=3.97, p=.063, but no WC1-BC effect, $F(1,17)$=.184, p=.674. There were no differences in the negativity evoked by WC1 and WC2 blanks in either of these anterior regions (all $F < .767$, all $p > .390$).
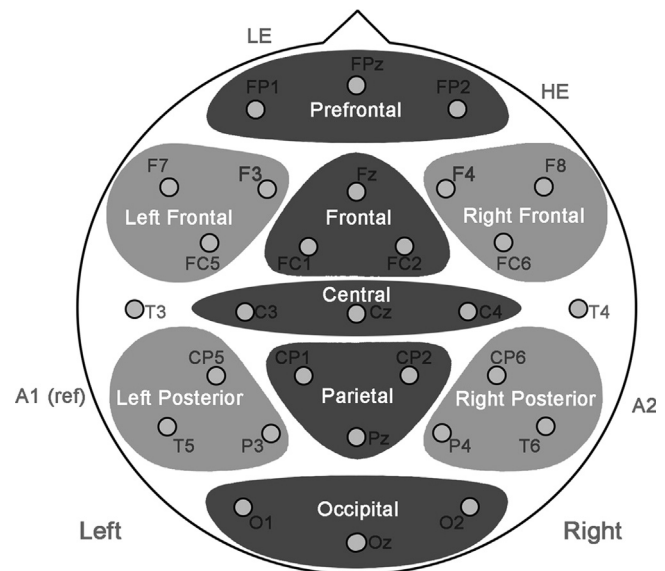

**Fig. 3.** Electrode montage analysis.

At the occipital region, the WC2 blank images evoked a larger positivity than BC blanks, $F(1,23)$=9.24, $p < .001$, but no such effect was seen in contrasting the WC1 and the BC blanks in any posterior regions (all $F < 1.82$, all $p > .190$). A direct contrast between the WC1 and WC2 blanks confirmed a marginally larger positivity evoked by the WC2 blanks, at the occipital region, $F(1,23)$=4.27, p=.05, and the right posterior region, $F(1,23)$=4.18, p=.05. A post-hoc analysis in the right posterior region indicated that this P600 effect began between 400 and 500 ms, but only for the WC2–WC1 effect, $F(1,17)$=10.6, $p < .01$.

4. Discussion

This experiment aimed to determine whether the brain draws upon a narrative constituent structure to comprehend sequences of visual images. To this end, we asked participants to view sequences of images, each conveying a simple narrative. We disrupted the structure of each narrative by inserting blank images within either its first or its second narrative constituent ("disruption panels"), and we compared the neural response produced by these structural violations to those produced by blank images falling at a natural constituent boundaries (between-constituent blanks). No ERP component patterned purely with the (average) linear position of the blank across the three conditions. Rather, relative to the between-constituent blanks, blank images that disrupted both the first and the second narrative constituent produced a left-lateralized anteriorly distributed negativity effect that was maximal between 500 and 700 ms. In addition, blank images that disrupted the second narrative constituent, but not the first, elicited a posteriorly-distributed positivity effect within the same time window.

4.1. The left anterior negativity effect

We suggest that the left anterior negativity effect reflects the neural response to a violation of comprehenders' fairly constrained anticipations of an upcoming narrative structure. On our interpretation, after viewing the first part of a narrative constituent, participants anticipated two or three panels that might possibly play particular narrative roles in this constituent. For example, in Fig. 1, following the Initial, comprehenders might anticipate a Peak, or possibly a Prolongation—a medial category separating Initials and
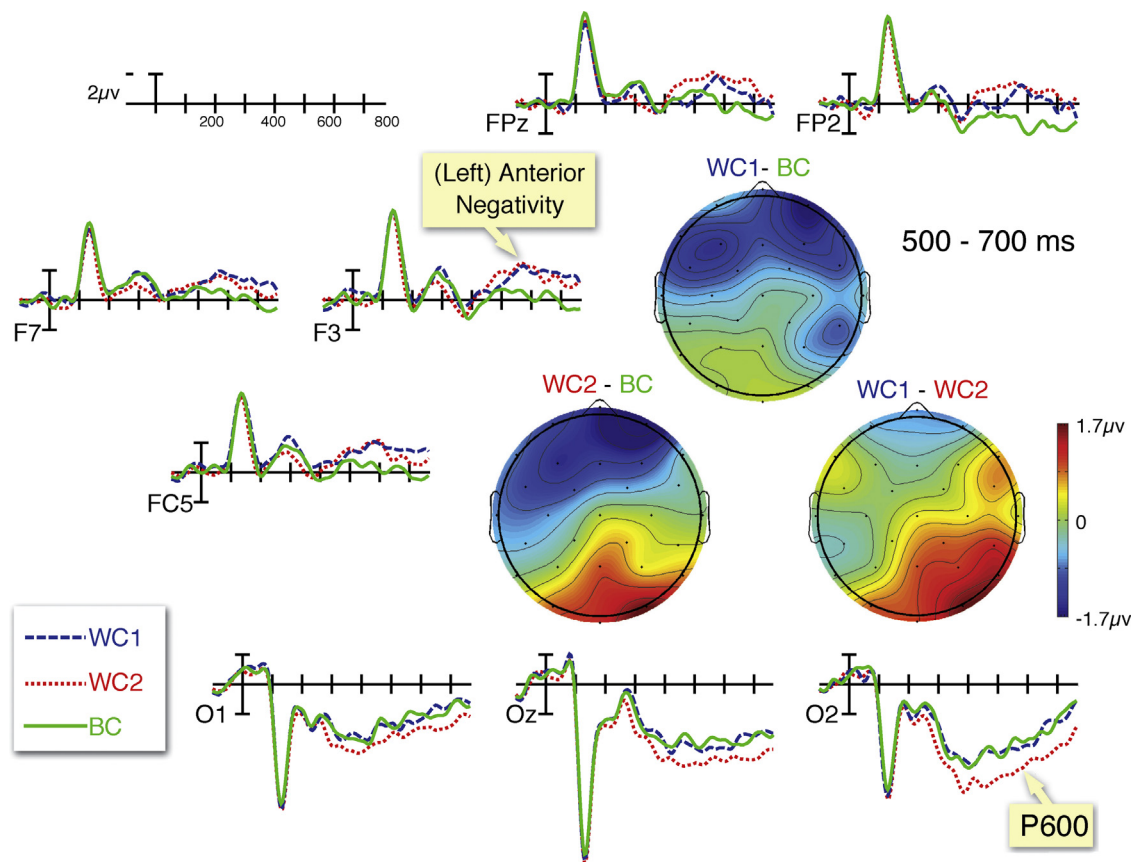
Fig. 4. ERP responses to blank images. ERP responses ($N=24$) time-locked to blank images that disrupted the first narrative constituent (WC1) or the second constituent (WC2), or that fell in between the two narrative constituents (BC). ERPs evoked by both the WC1 and WC2 blank images were more negative than those evoked by the BC blank images at frontal and leftward anterior sites (top and left). The WC2 blanks also evoked a more positive ERP waveform (a P600) than both the WC1 and BC blanks at posteriorly-distributed (bottom) electrode sites. Voltage maps illustrate the scalp distributions of these effects within the 500–700 ms time window.

Peaks (Cohn, 2014; Cohn & Paczynski, 2013). Upon encountering a blank image in this position (Fig. 2b), these expectations were violated, producing a relatively large neural response. In contrast, after viewing the entire first constituent (the Initial and the Peak in Fig. 1), participants predicted a natural constituent boundary. A blank image falling at this position, between the two narrative constituents (Fig. 2a), was more consistent with this prediction and produced a smaller neural response.

Importantly, this left anterior negativity effect cannot be attributed to the detection of one specific local violation (e.g., a blank always following an Initial) because we purposefully varied the structure of the sequences (the order/type of narrative constituents). For example, in some constituents, the blank in the first constituent followed an Initial, while in others it followed a Peak. Similarly, a blank between constituents might follow a Peak or a Release. In addition, the constituents varied in length (ranging from 2–4 panels long). This meant that a blank panel did not consistently occur at a single position in the sequence.[1]

A group of left anterior negativities have previously been associated with syntactic processing during language comprehension (Friederici, 2002; Friederici et al., 1993; Hagoort, Wassenaar, & Brown, 2003; Hahne & Friederici, 1999; Neville et al., 1991). The effect is particularly likely to be seen when a context constrains

for just two or three possible upcoming syntactic categories; its amplitude is smaller to input that matches (versus mismatches) one of these anticipations (Lau et al., 2006). Its onset latency can range from 100 ms (the so-called early left anterior negativity, ELAN e.g., Friederici et al., 1993; Hahne & Friederici, 1999; Neville et al., 1991), to 350 ms (the LAN, Hagoort et al., 2003). It has been suggested that this variation in onset latency reflects the speed with which the parser is able to determine whether the input matches or mismatches their syntactic predictions (Lau et al., 2006). This, in turn, will depend on cues in the input itself, including its morphological features (Hagoort et al., 2003) or even perceptual features that may be associated with the predicted syntactic category (Dikker, Rabagliati, Farmer, & Pylkkänen, 2010). In the present study, the left anterior negativity effect began relatively late, only becoming significant at 500 ms following the onset of the blank image. We suggest that this was because the blank panels did not provide obvious visual cues allowing comprehenders to make a fast diagnosis of structural match/mismatch. In this sense, the relatively delayed late anterior negativity effect in this study may be a more direct reflection of narrative or event structural surprisal/prediction error. This interpretation would be consistent with recent reports of late onset anterior negativities to words that match (versus mismatch) one of two possible event structures that are anticipated in certain sentence and discourse contexts (Baggio, van Lambalgen, & Hagoort, 2008; Paczynski, Jackendoff, & Kuperberg, 2014; Wittenberg, Paczynski, Wiese, Jackendoff, & Kuperberg, 2014; Wlotko & Federmeier, 2012).

The left lateralized distribution of the negativity effect seen to structural violations distinguishes it from another negative-going ERP component seen to both words and images—the N400

---

[1] This variation in structure meant that we could not determine whether the amplitude of the left anterior negativity was larger to blanks following some narrative categories versus others. If our interpretation is correct, however, then one might expect blanks following more predictive narrative categories (e.g., Peaks and Initials) to evoke a larger left anterior negativity than blanks following less predictive narrative categories.

between 300 and 500 ms, which is classically modulated by *semantic* congruity. The N400 effect evoked by language stimuli typically has a centro-parietal scalp distribution, while the N400 produced by images often has a wider scalp distribution and it is sometimes accompanied by a frontally-distributed (although not left lateralized) earlier negative-going peak (the N300) (Barrett & Rugg, 1990; Holcomb & McPherson, 1994). The N300/N400 complex is smaller to images containing semantically congruous versus incongruous elements (Mudrik, Lamy, & Deouell, 2010; Sitnikova, Holcomb, & Kuperberg, 2008b), to images preceded by semantically related (versus unrelated) "prime" images (Barrett & Rugg, 1990; Barrett et al., 1988), and to images preceded by semantically congruous (versus incongruous) narrative contexts (sequences of images or movies) (Cohn et al., 2012; Sitnikova et al., 2008b; Sitnikova, Kuperberg, & Holcomb, 2003; West & Holcomb, 2002).

The coherence of these narrative contexts is obviously influenced by their global narrative structure, and, in this sense, the N400 is *indirectly* sensitive to this structure. For example, in previous work we showed that the N300/N400 was smaller to target panels following coherent sequences (with both a global narrative structure and a common semantic theme) than to targets following sequences of panels that were related through a semantic theme but that did not have any narrative structure (Cohn et al., 2012). Importantly, however, the N300/N400 evoked by images is not *directly* sensitive to structural violations. For example, in our previous study, we saw no modulation of the N300/N400 complex when we contrasted sequences that were semantically incoherent without any narrative structure (i.e., fully scrambled sequences) to sequences that were semantically incoherent but that did have a narrative structure (i.e., structurally analogous to sentences like *Colorless green ideas sleep furiously*). Rather, this contrast was associated with a left anterior negativity effect with a similar distribution to that seen in the present study (although the effect in our previous study began earlier and lasted until around 900 ms). This finding provided preliminary evidence that left anterior negativity effects may be sensitive to structural processing during visual image comprehension. However, in this previous study, we did not explicitly violate the narrative structure of sequences and we did not target aspects of constituent structure directly; we only compared sequences in which the global narrative structure was present or absent. The present results therefore build upon these previous findings by showing that the left anterior negativity is the neural response produced when the input directly disconfirms comprehenders' anticipations of upcoming narrative constituent structure.

4.2. The P600 effect

Violations of the second narrative constituent, but not the first, produced a posteriorly distributed P600 effect in addition to the left anterior negativity effect. In previous work examining the comprehension of short, silent movie clips, we reported a similar P600 effect to images that violated strong predictions for specific events and event structures that were established by semantically constraining contexts, e.g., a man attempting to cut bread with an iron following a context showing him in the kitchen with the bread on a bread board (Sitnikova et al., 2003, 2008b). We suggested that the P600 was triggered by participants' detection of this strong event structure prediction violation, and that it reflected prolonged attempts to restructure and make sense of the input (Kuperberg, 2007; Sitnikova et al., 2008a, 2008b; see also Võ and Wolfe, 2013). In the present study, we offer a similar interpretation. Specifically, we argue that by the time comprehenders reached the second constituent, they had built a coherent semantic context and had generated strong, high certainty

predictions about both the semantic features and the narrative category of the upcoming image within this constituent. This is particularly likely given that comprehenders were asked explicitly to judge the coherence of each scenario. Encountering a blank image that disrupted this second narrative constituent violated these high certainty predictions and led to a large prediction error. It forced a grouping of the first three images together that comprehenders detected as being infelicitous (i.e., in Fig. 1, an ungrammatical constituent consisting of *Initial-Peak-Establisher*) and led to prolonged attempts to restructure the input in further efforts to make sense of the narrative as a whole. Within a generative predictive coding framework, this type of prediction error, leading to a P600, can be conceptualized as "unexpected surprise" (Kuperberg, 2013; Yu & Dayan, 2005) that led to a switch to (or learning of) a new internal model/latent cause that better explained this grouping of constituents.

Once again, there are analogies with language comprehension. There is a large literature describing P600 effects to violations of both syntactic and semantic structure within sentences (Hagoort et al., 1993; Kuperberg, 2007; Osterhout & Holcomb, 1992), as well as to violations of expected event structures in discourse (Ferretti, Rohde, Kehler, & Crutchley, 2009; Nieuwland & Van Berkum, 2005; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). The P600 evoked by linguistic violations is also triggered by input that violates a strong, high-certainty prediction for a specific semantic-structural mapping. Conflict between a strong, high certainty prediction and bottom-up input is most likely when the context is relatively rich and semantically constraining (Gunter et al., 2000; Nieuwland & Van Berkum, 2005) and when comprehenders carry out explicit coherence or acceptability judgment tasks (although neither is sufficient or necessary for evoking this effect, see Kuperberg, 2007). Such conflict is thought to lead to prolonged attempts to integrate both the structure and meaning of the input by updating or revising the wider representation of context (Christiansen, Conway, & Onnis, 2011; Kuperberg, 2013). Again, in Bayesian terms, this strong prediction violation can be conceptualized as unexpected surprise that triggers a switch to a new underlying model that might better explain the structural relationship between the context and the new input (Kuperberg, 2013; Yu & Dayan, 2005).[2]

4.3. General implications

Our findings are consistent with previous work showing that structure plays an important role in comprehending visual sequences. Previous studies have reported that participants tend to agree on the location of boundaries between episodes in visual narratives (Gernsbacher, 1985; Magliano & Zacks, 2011) and between visual events (Zacks et al., 2009; Zacks, Tversky, & Iyer, 2001). Recall of manipulated images in visual narratives also becomes less accurate when initial encoding entails crossing a

---

[2] There has been longstanding debate in the language literature about whether the P600 is a subtype of the well-known P300 (Coulson, King, & Kutas, 1998)—the ERP component that is triggered by violations of subjectively high probability predictions of a higher-order structure (Donchin & Coles, 1988; Wacongne et al., 2011), especially when these violations are task relevant. The P300 can also be evoked by subjectively unexpected omissions in structured sequences (Klinke et al., 1968; Simson et al., 1976; Wacongne et al., 2011), and it is thought to reflect the (re) allocation of attention to the context in order to update the model of the environment that gives rise to these predictions (Donchin & Coles, 1988). In the case of the P300, structural contextual predictions are established through sequences of relatively simple stimuli. During language comprehension, however, syntactic and semantic predictions are established through an interaction between the context and the stored rules and contingencies that constitute our linguistic knowledge. Similarly, in the present study, we argue that comprehender relied on his/her stored knowledge of narrative structure to generate structural expectations from the context.

constituent boundary (Carroll & Bever, 1976; Gernsbacher, 1985). Similarly, while viewing videos of real-world events, participants are more accurate in predicting subsequent actions within event segments than across event boundaries (Zacks, Kurby, Eisenberg, & Haroutunian, 2011). However, most of these previous approaches to studying visual events/narrative and narrative structure have attributed these segmentation effects to transient changes in semantics, such as shifts in characters or locations of an event following an event boundary (Gernsbacher, 1985; Magliano & Zacks, 2011). Because our results showed differences in the left anterior negativity between disruptions in WC1 and BC *prior* to the crossing of the boundary—i.e., before any shifts in characters or locations—they suggest that comprehenders build expectations image-by-image by using their knowledge of a more abstract narrative structure, rather than simply reacting to changes in semantic content. This interpretation is consistent with previous work that has reported anticipatory neural activity prior to expected event boundaries during non-verbal comprehension (Kurby & Zacks, 2008; Zacks et al., 2001).

Finally, it is important to note that visual narrative is not the first domain argued to have some sort of constituent structure analogous to that of language. "Grammatical" systems have been proposed in vision (Biederman, 1987; Marr, 1982), drawing (Cohn, 2012; Willats, 2005), social relations (Jackendoff, 2007), and music (Jackendoff, 2011; Lerdahl & Jackendoff, 1982). Like the ERPs elicited by manipulations to visual narratives in this study, violations of musical structure have also evoked waveforms similar to those in language processing: posteriorly distributed P600 effects and anterior negativities—in these cases, often with a lateralized rightward distribution (Koelsch, Gunter, Wittfoth, & Sammler, 2005; Koelsch & Siebel, 2005; Patel, 2003; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). This of course does not imply that the structures that we draw upon to make sense of language, music, and visual narrative are the same; each of these cognitive domains differ both in their basic units (words, notes, and images) and in the rules by which they are combined (syntax, harmony, and narrative). What it does suggest, however, is that constituent structure is not specific to language, and that the brain draws upon similar neurocognitive mechanisms or common computational principles to analyze structure across multiple domains (Corballis, 1991; Hoen & Dominey, 2000; Jackendoff, 2011; Patel, 2003; Sitnikova et al., 2008a; Võ & Wolfe, 2013).

Acknowledgments

References

Abrams, K., Bever, T.G., 1969. Syntactic structure modifies attention during speech perception and recognition. Quarterly Journal of Experimental Psychology 21 (3), 280–290. http://dx.doi.org/10.1080/14640746908400223.

Baggio, G., van Lambalgen, M., Hagoort, P., 2008. Computing and recomputing discourse models: an ERP study. Journal of Memory and Language 59 (1), 36–53.

Baird, J.A., Baldwin, D.A., 2001. Making sense of human behavior: action parsing and intentional inference. In: Malle, B.F., Moses, L.J., Baldwin, D.A. (Eds.), Intention and intentionality. MIT Press, Cambridge.

Barrett, S.E., Rugg, M.D., 1990. Event-related potentials and the semantic matching of pictures. Brain and Cognition 14 (2), 201–212.

Barrett, S.E., Rugg, M.D., Perrett, D.I., 1988. Event-related potentials and the matching of familiar and unfamiliar faces. Neuropsychologia 26 (1), 105–117.

Bendixen, A., Schröger, E., Winkler, I., 2009. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. The Journal of Neuroscience 29 (26), 8447–8451. http://dx.doi.org/10.1523/jneurosci.1493-09.2009.

Berent, I., Perfetti, C.A., 1993. An on-line method in studying music parsing. Cognition 46 (3), 203–222. http://dx.doi.org/10.1016/0010-0277(93)90010-s.

Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. Psychological Review 94, 115–147.

Bond, Z.S., 1972. Phonological units in sentence perception. Phonetica 25 (3), 129–139.

Butcher, S.H., 1902. The poetics of Aristotle, 3rd ed. Macmillian and Co. Ltd, London.

Carroll, J.M., Bever, T.G., 1976. Segmentation in cinema perception. Science 191 (4231), 1053–1055.

Chomsky, N., 1965. Aspects of the theory of syntax. MIT Press, Cambridge, MA.

Christiansen, M.H., Conway, C.M., Onnis, L., 2011. Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. Language and Cognitive Processes 27 (2), 231–256. http://dx.doi.org/10.1080/01690965.2011.606666.

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences 36 (03), 181–204. http://dx.doi.org/10.1017/S0140525 × 12000477.

Clark, H.H., 1996. Using language. Cambridge University Press, Cambridge, UK.

Cohn, N., 2012. Explaining "I can't draw": parallels between the structure and development of language and drawing. Human Development 55 (4), 167–192. http://dx.doi.org/10.1159/000341842.

Cohn, N., 2013a. The visual language of comics: introduction to the structure and cognition of sequential images. Bloomsbury, London, UK.

Cohn, N., 2013b. Visual narrative structure. Cognitive Science 37 (3), 413–452 http://dx.doi.org/10.1111/cogs.12016.

Cohn, N., 2014. You're a good structure, Charlie Brown: the distribution of narrative categories in comic strips. Cognitive Science 38 (7), 1317–1359. http://dx.doi.org/10.111/cogs.12116.

Cohn, N., Paczynski, M., 2013. Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. Cognitive Psychology 67 (3), 73–97.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P.J., Kuperberg, G.R., 2012. (Pea)nuts and bolts of visual narrative: structure and meaning in sequential image comprehension. Cognitive Psychology 65 (1), 1–38. http://dx.doi.org/10.1016/j.cogpsych.2012.01.003.

Corballis, M., 1991. The lopsided ape: evolution of the generative mind. Oxford University Press, Oxford.

Coulson, S., King, J., Kutas, M., 1998. Expect the unexpected: event-related brain responses to morphosyntactic violations. Language and Cognitive Processes 13 (71–74), 71.

Culicover, P.W., Jackendoff, R., 2005. Simpler syntax. Oxford University Press, Oxford.

Dikker, S., Rabagliati, H., Farmer, T.A., Pylkkänen, L., 2010. Early occipital sensitivity to syntactic category is based on form typicality. Psychological Science 21 (5), 629–634. http://dx.doi.org/10.1177/0956797610367751.

Donchin, E., Coles, M.G. H., 1988. Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences 11 (03), 357–374. http://dx.doi.org/10.1017/S0140525 × 00058027.

Ferretti, T.R., Rohde, H., Kehler, A., Crutchley, M., 2009. Verb aspect, event structure, and coreferential processing. Journal of Memory and Language 61 (2), 191–205.

Fodor, J., Bever, T.G., 1965. The psychological reality of linguistic segments. Journal of Verbal Learning and Verbal Behavior 4 (5), 414–420.

Ford, M., Holmes, V.M., 1978. Planning units and syntax in sentence production. Cognition 6 (1), 35–53. http://dx.doi.org/10.1016/0010-0277(78)90008-2.

Freytag, G., 1894. Technique of the drama. S.C. Griggs & Company, Chicago.

Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences 6 (2), 78–84.

Friederici, A.D., Pfeifer, E., Hahne, A., 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Cognitive Brain Research 1 (3), 183–192.

Friston, K., 2005. A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1456), 815–836. http://dx.doi.org/10.1098/rstb.2005.1622.

Garrett, M.F., Bever, T.G., 1974. The perceptual segmentation of sentences. In: Bever, T.G., Weksel, W. (Eds.), The structure and psychology of language. Mouton and Co., The Hague.

Gernsbacher, M.A., 1985. Surface information loss in comprehension. Cognitive Psychology 17, 324–363.

Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. Journal of Cognitive Neuroscience 12, 556–568.

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, C.M., Groothusen, J., 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. In: Garnsey, S.M. (Ed.), Language and cognitive processes. Special issue: Event-related brain potentials in the study of language, 8. Lawrence Erlbaum Associates, Hove, pp. 439–483.

Hagoort, P., Wassenaar, M., Brown, C.M., 2003. Syntax-related ERP-effects in Dutch. Cognitive Brain Research 16 (1), 38–50.

Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. Journal of Cognitive Neuroscience 11 (2), 194–205. http://dx.doi.org/10.1162/089892999563328.

Hinds, J., 1976. Aspects of Japanese discourse. Kaitakusha Co., Ltd, Tokyo.

Hoen, M., Dominey, P.F., 2000. ERP analysis of cognitive sequencing: a left anterior negativity related to structural transformation processing. NeuroReport 11 (14), 3187–3191.

Holcomb, P., McPherson, W.B., 1994. Event-related brain potentials reflect semantic priming in an object decision task. Brain and Cognition 24, 259–276.

Jackendoff, R., 2007. Language, consciousness, culture: essays on mental structure (Jean Nicod Lectures). MIT Press, Cambridge, MA.

Jackendoff, R., 2011. What is the human language faculty? Two views. Language 87 (3), 586–624.

Klinke, R., Fruhstorfer, H., Finkenzeller, P., 1968. Evoked responses as a function of external and stored information. Electroencephalography and Clinical Neurophysiology 25 (2), 119–122.

Koelsch, S., Gunter, T.C., Wittfoth, M., Sammler, D., 2005. Interaction between syntax processing in language and in music: an ERP study. Journal of Cognitive Neuroscience 17 (10), 1565–1577.

Koelsch, S., Siebel, W.A., 2005. Towards a neural basis of music perception. Trends in Cognitive Sciences 9 (12), 578–584. http://dx.doi.org/10.1016/j.tics.2005.10.001.

Kung, S.-J., Tzeng, O., Hung, D., Wu, D., 2011. Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences. Experimental Brain Research 210 (2), 269–282. http://dx.doi.org/10.1007/s00221-011-2630-2.

Kunzle, D., 1973. The history of the comic strip, 1. University of California Press, Berkeley.

Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Research 1146, 23–49.

Kuperberg, G., 2013. The pro-active comprehender: what event-related potentials tell us about the dynamics of reading comprehension. In: Miller, B., Cutting, L., McCardle, P. (Eds.), Unraveling the behavioral, neurobiological, and genetic components of reading comprehension. Paul Brookes Publishing, Baltimore.

Kurby, C.A., Zacks, J.M., 2008. Segmentation in the perception and memory of events. Trends in Cognitive Science 12 (2), 72–79.

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: Brain potential reflect semantic incongruity. Science 207, 203–205.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology 62 (1), 621–647.

Labov, W., Waletzky, J., 1967. Narrative analysis: oral versions of personal experience. In: Helm, J. (Ed.), Essays on the verbal and visual arts. University of Washington Press, Seattle, pp. 12–44.

Lau, E., Stroud, C., Plesch, S., Phillips, C., 2006. The role of structural prediction in rapid syntactic analysis. Brain and Language 98 (1), 74–88.

Lerdahl, F., Jackendoff, R., 1982. A generative theory of tonal music. MIT Press, Cambridge, MA.

Magliano, J.P., Zacks, J.M., 2011. The impact of continuity editing in narrative film on event segmentation. Cognitive Science 35 (8), 1489–1517.

Mandler, J.M., Johnson, N.S., 1977. Remembrance of things parsed: story structure and recall. Cognitive Psychology 9, 111–151.

Marr, D., 1982. Vision. Freeman, San Francisco, CA.

McCloud, S., 1993. Understanding comics: the invisible art. Harper Collins, New York, NY.

Mudrik, L., Lamy, D., Deouell, L.Y., 2010. ERP evidence for context congruity effects during simultaneous object-scene processing. Neuropsychologia 48 (2), 507–517.

Münte, T.F., Matzke, M., Johannes, S., 1997. Brain activity associated with syntactic incongruencies in words and psuedo-words. Journal of Cognitive Neuroscience 9, 318–329.

Neville, H.J., Nicol, J.L., Barss, A., Forster, K.I., Garrett, M.F., 1991. Syntactically based sentence processing classes: evidence from event-related brain potentials. Journal of Cognitive Neuroscience 3 (2), 151–165.

Nieuwland, M.S., Van Berkum, J.J. A., 2005. Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. Cognitive Brain Research 24 (3), 691–701. http://dx.doi.org/10.1016/j.cogbrainres.2005.04.003.

Osterhout, L., Holcomb, P., 1992. Event-related potentials elicited by syntactic anomaly. Journal of Memory and Language 31, 758–806.

Paczynski, M., Jackendoff, R., Kuperberg, G., 2014. When events change their nature: the neurocognitive mechanisms underlying aspectual coercion. Journal

of Cognitive Neuroscience 26 (9), 1905–1917. http://dx.doi.org/10.1162/jocn_a_00638.

Patel, A.D., 2003. Language, music, syntax and the brain. Nature Neuroscience 6 (7), 674–681.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., Holcomb, P.J., 1998. Processing syntactic relations in language and music: an event-related potential study. Journal of Cognitive Neuroscience 10 (6), 717–733.

Rao, R.P. N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2 (1), 79–87.

Rumelhart, D.E., 1975. Notes on a schema for stories. In: Bobrow, D., Collins, A. (Eds.), Representation and understanding. Academic Press, New York, NY, pp. 211–236.

Schank, R.C., Abelson, R., 1977. Scripts, plans, goals and understanding. Lawrence Earlbaum Associates, Hillsdale, NJ.

Simson, R., Vaughan Jr., H.G., Walter, R., 1976. The scalp topography of potentials associated with missing visual or auditory stimuli. Electroencephalography and Clinical Neurophysiology 40 (1), 33–42.

Sitnikova, T., Holcomb, P.J., Kuperberg, G.R., 2008a. Neurocognitive mechanisms of human comprehension. In: Shipley, T.F., Zacks, J.M. (Eds.), Understanding events: how humans see, represent, and act on events. Oxford University Press, Oxford, UK, pp. 639–683.

Sitnikova, T., Holcomb, P.J., Kuperberg, G.R., 2008b. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. Journal of Cognitive Neuroscience 20 (11), 1–21.

Sitnikova, T., Kuperberg, G.R., Holcomb, P., 2003. Semantic integration in videos of real-world events: an electrophysiological investigation. Psychophysiology 40 (1), 160–164.

Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. The Journal of Neuroscience 31 (25), 9118–9123. http://dx.doi.org/10.1523/jneurosci.1425-11.2011.

Van Berkum, J.J. A., Koornneef, A.W., Otten, M., Nieuwland, M.S., 2007. Establishing reference in language comprehension: An electrophysiological perspective. Brain Research 1146 (0), 158–171.

Võ, M.L.-H., Wolfe, J.M., 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. Psychological Science 24 (9), 1816–1823.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., Dehaene, S., 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. Proceedings of the National Academy of Sciences of the United States of America 108 (51), 20754–20759. http://dx.doi.org/10.1073/pnas.1117807108.

West, W.C., Holcomb, P., 2002. Event-related potentials during discourse-level semantic integration of complex pictures. Cognitive Brain Research 13, 363–375.

Willats, J., 2005. Making sense of children's drawings. Lawrence Erlbaum, Mahwah, NJ.

Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., Kuperberg, G., 2014. The difference between "giving a rose" and "giving a kiss": sustained neural activity to the light verb construction. Journal of Memory and Language 73 (0), 31–42.

Wlotko, E.W., Federmeier, K.D., 2012. So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. NeuroImage 62 (1), 356–366.

Yu, A.J., Dayan, P., 2005. Uncertainty, neuromodulation, and attention. Neuron 46 (4), 681–692. http://dx.doi.org/10.1016/j.neuron.2005.04.026.

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., et al., 2001. Human brain activity time-locked to perceptual event boundaries. Nature Neuroscience 4 (6), 651–655.

Zacks, J.M., Kurby, C.A., Eisenberg, M.L., Haroutunian, N., 2011. Prediction error associated with the perceptual segmentation of naturalistic events. Journal of Cognitive Neuroscience 23 (12), 4057–4066. http://dx.doi.org/10.1162/jocn_a_00078.

Zacks, J.M., Speer, N.K., Reynolds, J.R., 2009. Segmentation in reading and film comprehension. Journal of Experimental Psychology: General 138 (2), 307–327.

Zacks, J.M., Tversky, B., Iyer, G., 2001. Perceiving, remembering, and communicating structure in events. Journal of Experimental Psychology 130 (1), 29–58.