Segmenting Visual Narratives: Evidence for Constituent Structure in Comics

Neil Cohn (neilcohn@emaki.net), Phillip Holcomb
Department of Psychology, 490 Boston Ave
Medford, MA 02155 USA

Ray Jackendoff
Center for Cognitive Studies, Miner Hall
Medford, MA 02155 USA

Gina Kuperberg
Department of Psychology, 490 Boston Ave
Medford, MA 02155 USA
Department of Psychiatry and Athinoula A. Martinos Center for Biomedical Imaging
Massachusetts General Hospital, Bldg 149, 13th Street
Charlestown, MA 02129 USA

Abstract

We have proposed that visual narratives in comics are organized with a hierarchic narrative "grammar." Inspired by classic "click" studies of syntax, we inserted blank "disruption" panels Before, At, or After the constituent boundaries of comic strips. In self-paced viewing, Experiment 1 found that blanks After the boundary were viewed slower than Before or At the boundary. Panels immediately following blanks were slower than corresponding panels in sequences without blank panels, but only when placed Before or After the boundary. Three ordinal panel positions following the boundary, panels following blanks At the boundary or with No-Blanks were viewed faster than panels following blanks After or Before the boundary. This supports constituency because disruptions had greater impact within, as opposed to between, constituents. Using ERPs, Experiment 2 found a larger anterior negativity to blanks within constituents (Before/After) than between constituents (At). This indicates disruptions of constituents are recognized before reaching a subsequent panel. A larger P600 appeared to blanks After the boundary than in the first constituent (Before/At). This positivity may reflect a reanalysis reflecting the inability to integrate all prior panels into a single constituent, since they are divided by the constituent boundary.

Keywords: Grammar; Constituent Structure; Discourse; Narrative; Comics; ERPs; Left Anterior Negativity.

Introduction

Drawings have conveyed narratives through sequences of images for thousands of years, but in contemporary society they appear most prevalently in comics. In comparison with research on the structure and comprehension of verbal narrative, however, little is known about mechanisms of processing sequential images. For example, are narrative units integrated linearly across each adjacent relationship or they organized into hierarchic constituents?

Background

Since the 1950s, research on language has stressed that sentences are organized into hierarchic constituents, rather than linear word-to-word connections (e.g., Chomsky, 1957).

An analogous distinction between local relationships and hierarchic segmentation has underlined research of text and discourse. Traditional approaches emphasize coherence relationships between individual sentences (Halliday & Hasan, 1976; Zwaan & Radvansky, 1998). Participants can intuitively segment texts into consistent groupings (Gee & Grosjean, 1984; Mandler & Johnson, 1977), suggesting that readers' comprehension extends beyond adjacent relationships between individual sentences.

Some theories have formally described hierarchical relationships in narrative, such as "story grammars" (e.g. Mandler & Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977), which use phrase structure rules to organize narratives around characters' goal-driven events. Constituent structure in these models was examined using clustering models (Gee & Grosjean, 1984; Mandler, 1987; Mandler & Johnson, 1977), similar to those employed by early psycholinguistic research on syntactic relations (e.g. Levelt, 1970). Participants divided a narrative into logical groupings, and these intuitions were then submitted to the algorithms in hierarchic clustering models. These analyses yielded tree structures that closely correlated with the models predicted by the original theories.

As in verbal discourse, theories of visual narrative have also emphasized linear relationships between individual images (McCloud, 1993). However, beyond such linear relationships, participants also intuitively divide visual narratives into segments (Gernsbacher, 1985). Again, this suggests that comprehension may extend beyond adjacent relationships.

Recently, Cohn (2003, In Press) has proposed a theoretical model of narrative structure that formalizes the

constituent structure of visual narrative. Like story grammars' treatment of sentences, this model organizes panels into hierarchic constituents (though important differences distinguish this approach from story grammars, and this model could potentially provide an alternative approach for describing the structure of verbal discourse).

Figure 1 illustrates the narrative structure for a 6-panel *Peanuts* comic strip. This sequence depicts a baseball game that starts with Lucy hitting a ball, which allows Charlie Brown to run home and score, escaping a tag by Schroeder. The first panel shows Lucy tossing a ball. This panel functions as an Initial, initiating the interactions in the sequence. In the second panel, Lucy hits the ball, a narrative "Peak" as the culmination of the initiated action. Together, these two panels act as an Initial constituent that propels the rest of the sequence. A second constituent begins with Schroeder waiting for the ball—nothing happens here except a set-up of the characters (Establisher). The next Initial begins the new set of events, which climax in the penultimate Peak panel—Charlie interrupts the catch by sliding into the base. The last panel then resolves this interaction (Release). Together, these panels act as a Peak that is set up by the constituent-level Initial of the first panels. Thus, at a higher level of processing, the first constituent (Lucy hitting the ball) acts as an Initial, which facilitates the second constituent (Charlie scoring), a Peak. As a result, the narrative structure operates on both the panel level and the level of whole constituencies.
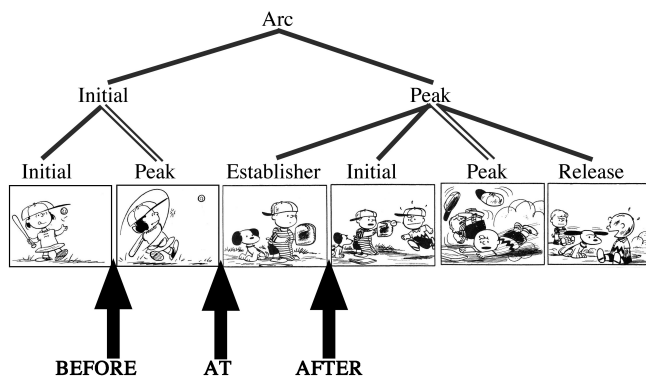

Figure 1: Narrative structure in a comic strip with markings Before, At, and After the narrative constituent boundary.

While clustering analyses have shown evidence for hierarchic structure in verbal discourse, no studies to date have examined constituent in sequential images. Furthermore, evidence that narrative constituent structure is used during *online* processing has yet to be explored in any modality. Here we describe two experiments to determine whether constituents are used when processing sequential images.

Experiment 1: Self-paced Viewing

In a classic experiment on syntactic structure, Fodor and Bever (1965) pioneered a "click technique" where they played a verbal sentence in one ear of a participant, and then introduced short bursts of white noise ("clicks") in the other ear. They reasoned that, if clauses constitute the perceptual processing units of sentences, clicks that disrupt those units would be harder to discern than clicks occurring between clauses. They found that participants were better able to recall clicks that were placed at the clause boundaries than those before or after it. Also, participants tended to falsely recall clicks within constituencies as occurring within the constituency break. Subsequent studies found similar findings, with the overall interpretation that such disruptions reflect the psychological validity of a syntactic constituent structure (for review see Garrett & Bever, 1974).

In Experiment 1, we used an analogous "disruption" paradigm to determine whether the comprehension of visual sequences also draws upon a constituent structure in narrative sequences. We measured viewing times in graphic sequences where blank white "disruption" panels were inserted Before, At, or After the narrative constituency boundary.

Methods

Participants
60 self-defined comic readers (35 male, 27 female, mean age = 24.03) from the Tufts University student population and surrounding neighborhoods were paid for their participation.

Stimuli
Novel 6-frame long sequences were created (160 sets) using individual panels culled from several volumes of Charles Schulz's *Peanuts*. Using Cohn's (In Press) model of visual narrative, we designed sequences that had two narrative constituents. This was confirmed using a behavioral task in which 20 participants drew lines to divide strips into two parts. Our final 120 strips had a 71% agreement on where constituent boundaries were located. Each sequence had constituent boundaries appearing after panel 2, 3, or 4 (40 of each type). Using these strips, we then inserted blank "disruption" panels Before, At, or After the constituency boundaries (as notated at the bottom of Figure 1), along with No-Blank control sequences. Because of the variation in position of the boundary, blank panels across all sequences could appear between the second and fifth panel positions, and items were counterbalanced across four lists such that each participant only viewed a sequence once. 15 fillers in each list had two successive blank panels and one-third of 75 additional no-blank fillers had violations of coherence.

Procedure

Participants viewed comic panels one frame at a time on a computer screen, self-pacing their way through the sequence. After each sequence, they rated the coherence of the strip on a 1 to 5 scale.

Results

Three-way ANOVAs showed that the position of the blank panel significantly impacted how fast it was viewed (see Figure 1), $F1(2,118)=12.93$, $p<.005$, $F2(2,238)=7.26$, $p<.005$. This effect arose because blanks After the boundary were viewed significantly slower than blanks Before and At the boundary, (all ts < -2.7, all ps $< .01$). There were no difference in viewing times between blanks appearing Before and At the boundary, $t1(59)=-.964$, $p=.339$, $t2(119)=-1.09$, $p=.280$.
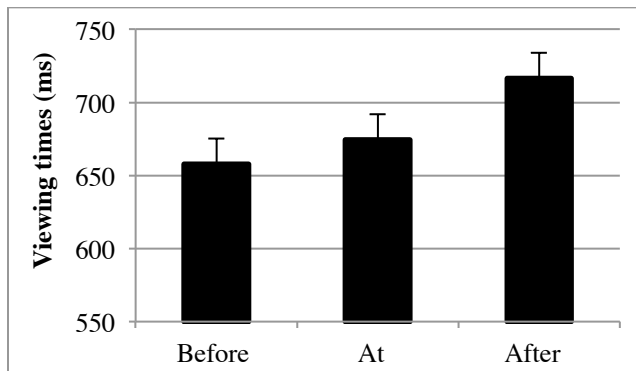

Figure 2: Viewing times to blank panels.

The placement of blank panels in a sequence also impacted the processing of subsequent panels. Viewing times to panels appearing immediately after the blank differed from corresponding panels in sequences that had No-Blank (i.e., Blank +1, see Figure 3). Overall, we found a main effect of Disruption (blank, no blank), $F1(1,59)=6.34$, $p<.05$, $F2(1,119)=14.11$, $p<.001$, a main effect of Position (Before, At, After) in the subjects analysis, $F1(2,118)=3.11$, $p<.05$, $F2(2,238)=.472$, $p=.617$, but no interaction between them, $F1(2,118)=1.3$, $p=.276$, $F2(2,238)=1.57$, $p=.210$. Panels following blanks Before the narrative constituent boundary were viewed significantly slower than their corresponding No-Blank panels, $t1(59)=-3.11$, $p<.005$, $t2(119)=-3.48$, $p<.005$, as were panels following blanks After the boundary, a trending difference in the items analysis, $t1(59)=-1.34$, $p=.184$, $t2(119)=-1.8$, $p=.074$. Panels following blanks At the boundary were not slower than corresponding panels in No-Blank sequences, $t1(59)=-.878$, $p=.383$, $t2(119)=-1.56$, $p=.121$.
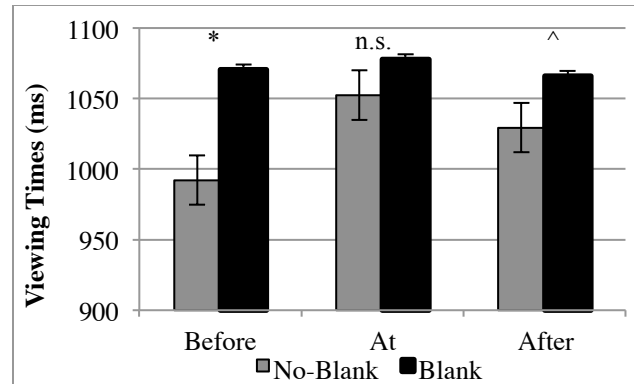

Figure 3: Viewing times to panels immediately after the blank panel (Blank +1) compared with corresponding panels in sequences with No-Blanks.

Delayed effects were also found three panel positions after the narrative constituent boundary (Boundary+3, see Figure 4), $F1(3,177)=2.90$, $p<.05$, $F2(3,357)=5.24$, $p=.005$. Panels following a blank After the boundary were viewed slower than panels following blanks At the boundary and panels in sequences with No-Blank, (all ts > -3.36, all ps $< .05$). Panels following blanks Before the boundary were viewed slower than those following blanks At the boundary, $t1(59)=1.49$, $p=.142$, $t2(79)=2.12$, $p<.05$, and trending to be slower than panels in sequences with No-Blanks, $t1(59)=-1.56$, $p=.124$, $t2(79)=-1.77$, $p=.081$. However, viewing times did not differ between panels following blanks Before and After the boundary, or between panels following a blanks At the boundary and in sequences with No-Blank, (all ts < -1.34, p $> .184$).
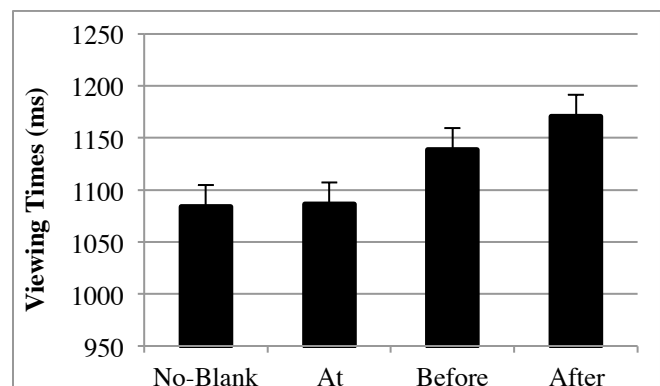

Figure 4: Viewing times to panels three positions after the narrative constituent boundary (Boundary +3) in all sequence types.

Discussion

Blank panels were viewed slower when they appeared After compared to Before or At the narrative constituent boundary. Panels immediately following blanks were slower than corresponding panels in sequences without a preceding

blank panel, but only when placed Before or After the boundary. A delayed effect of this disruption appeared three panels after the narrative constituent boundary. At this position, panels following blanks At the boundary or with No-Blanks were viewed faster than panels following blanks After or Before the boundary. These results are consistent with the presence of constituent structure to visual narrative: a blank panel disruption had greater impact within (Before/After) as opposed to between (At) constituents.

Experiment 2: Event-Related Potentials

Different ERP components have been associated with manipulations of semantic and syntactic constraints. For example, the N400 ERP component has been associated with semantic processing of words (Kutas & Hillyard, 1980), individual visual images (Holcomb & McPherson, 1994), and images in visual narratives images (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; West & Holcomb, 2002). The N400 effect has been suggested as indexing the spreading activation of a word or image's meaning as it integrates with its preceding context to the information stored in semantic memory (Kutas & Federmeier, 2011).

In contrast, two ERP components have been linked to violations of grammatical structure during sentence processing. First, the P600 is a positive deflection that peaks from 600-800ms (Osterhout & Holcomb, 1992). It is seen to syntactic anomalies and ambiguities and has been interpreted as reflecting a process of continued analysis and/or repair as structural and semantic information are integrated to make sense of a sentence (Friederici, 2002; Kuperberg, 2007).

Waveforms resembling the P600 have also been found in domains outside of language. For example, in studies using silent movie clips of everyday events, a P600 was seen to "action violation" endings in which a predicted action was carried out with an incongruous object (e.g. a person preparing to cut a piece of bread followed by an image of the person attempting to cut the bread with an iron). This suggested that the P600 effect may reflect the integration of structural and semantic information around an event beyond linguistic processing (Sitnikova, Holcomb, & Kuperberg, 2008).

Certain syntactic operations have also been tied to a left-anterior negativity (LAN), which falls in the same time window as the N400 (between 300 and 500ms), but is distributed over frontal and left lateralized regions (Neville, Nicol, Barss, Forster, & Garrett, 1991). This component has been associated with violations of syntactic constituent structure in sentences (Friederici, 2002).

Anterior negativities outside of language have also been associated with violations of hierarchic structure in music. Patel and colleagues (1998) found a P600 to structural violations in musical sequences (e.g., a nearby key chord or a distant-key chord appearing after an otherwise in-key musical sequence). Another negative-going effect, distributed over right anterior and temporal sites, also

appeared between 300 and 400 milliseconds. This (*early*) *right anterior negativity* has led researchers to argue for overlap in the neural resources used to process structure in both music and language (Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, et al., 1998).

Only one study has previously hinted at a LAN during the processing of visual sequences. Cohn et al. (2012) manipulated the presence or absence of narrative structure and semantic relatedness between panels in visual narratives. A larger negativity was found to sequences of scrambled panels (which had no semantic relations and no narrative structure) than to sequences with only narrative structure but no semantic theme (i.e., the sequence followed a narrative arc, but used panels from various sequences which had no meaningful relationship with each other). However, this negativity only appeared in a localized left anterior region, and was larger in participants with higher comic reading expertise. This left anterior effect was distinguished from a more widespread N400 effect that was greater in magnitude in these sequences that lacked semantic relations than to panels in sequences with no narrative structure, but with semantic relations to a general theme (like baseball), and to semantically and narratively congruous Normal sequences. We suggested that the left localized negativity—distinguished form the N400—might be analogous to the left anterior negativity effect seen to violations of structure in language. However, because this study did not directly introduce violations to narrative structure, this interpretation was speculative.

In Experiment 2, we predicted that violations of constituency in visual narratives would elicit similar ERP effects as violations of structure during sentence processing: P600 and LAN effects. We used the same set of stimuli from Experiment 1 and measured ERPs directly at blank panels.

Methods

Participants
24 self-defined comic readers from Tufts University (8 male, 16 female, mean age = 19.9) participated in the study for compensation.

Stimuli
The same stimuli were used in Experiment 2 as in Experiment 1.

Procedure
Participants viewed each sequence one panel at a time on a computer screen while ERPs were measured to all panels. After each sequence, participants rated its coherence on a 1 to 5 scale.

Results
At the blank panel, between 500 and 700 milliseconds, repeated measures ANOVAs revealed significant interactions between Disruption Position, Region, and

Electrode Site at midline regions, and interactions between Disruption Position, Region, and Hemisphere at peripheral regions of the scalp (all Fs > 3.24, all ps < .05). Follow-ups showed a larger negativity to blank panels that appeared Before the boundary than to those appearing At the boundary at left anterior regions (F3, FC5, F7). Blanks appearing After the boundary evoked a larger negativity than those At the boundary again in left anterior regions (F3, FC5, F7) as well as at FPz and FP2. ERPs to Blank panels appearing Before or After the boundary did not differ in amplitude.
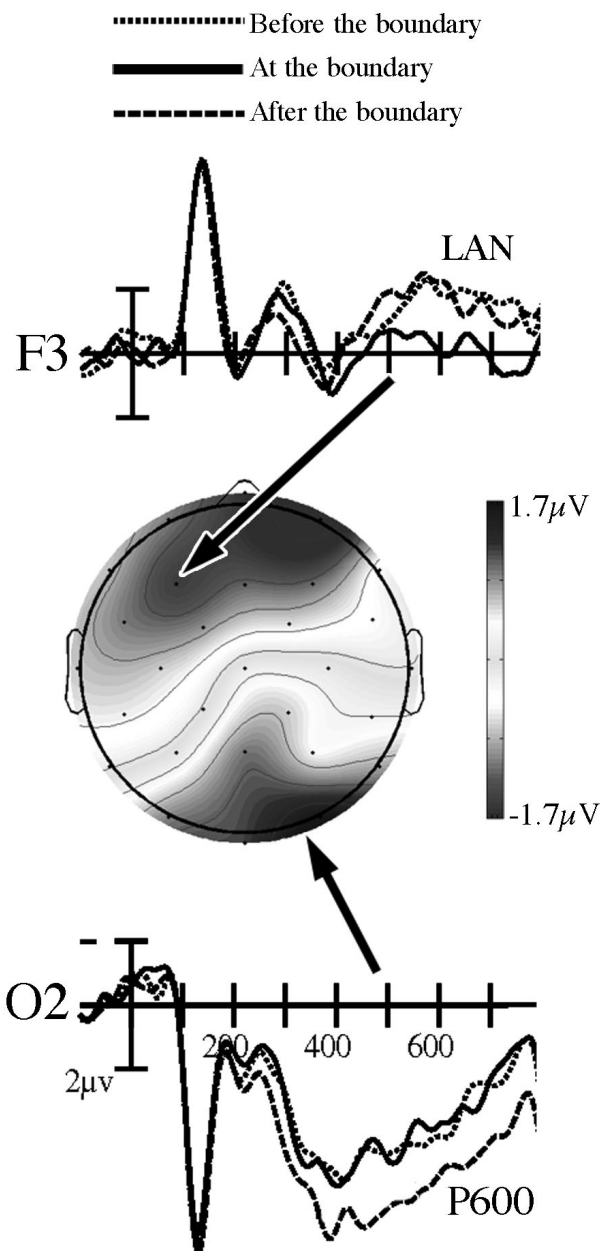

Figure 5: Amplitudes for blank panels at left anterior (F3) and posterior (O2) sites. Scalp map shows the contrast between blank panels After and At the boundary from 500 and 700ms.

At posterior sites, a larger P600 appeared to blank panels placed After the boundary (within the second constituent) than those placed Before or At the constituency boundary. The contrast between the After and At conditions revealed a significant effect at O1, Oz, and O2, and between the After and Before conditions showed an effect at O2.

Discussion

A larger negativity was found to disruption blank panels that appeared within constituents (Before/After) than to blanks that appeared between constituents (At). This negativity localized to anterior, and somewhat left lateralized, locations. This is consistent with the lateralized anterior distribution of negativity effects shown to violations of syntax in sentences (Neville, et al., 1991) and music (Koelsch, et al., 2005; Patel, et al., 1998). It is important to note that it is only possible to confirm or verify that a blank panel has actually disrupted a narrative constituent once the subsequent panel is reached. However, we still saw a larger anterior negativity where the disruption occurred Before than At the boundary on the blank panel itself, prior to confirmation on the subsequent panel. This suggests that the brain may make online predictions about the building of constituent structure as a narrative progresses.

In addition to the anterior negativity effect, we also saw a P600 effect to blank panels appearing After (versus Before or At) the boundary. Because this positivity appeared to blanks only after a new constituent had been reached, we suggest that it reflected the failure of integrating all of the prior panels into a single constituent. This blank followed a panel after the constituent boundary, meaning that the preceding panel would be unable to be integrated into a single constituent with the other prior panels, thereby evoking a reanalysis of starting a new constituent (Friederici, 2002; Kuperberg, 2007).

General Discussion

The results of Experiment 1 suggest that the introduction of a blank panel into a sequence has a greater impact on viewing times of subsequent panels if it falls within a narrative constituent (Before/After a narrative boundary) than between narrative constituents (At a narrative boundary). This provides evidence that comprehenders use narrative boundaries during the processing of visual sequences. The results of Experiment 2 support this interpretation. Moreover, because effects were seen on the blank panel itself, this further suggests that comprehenders of visual sequences make active predictions about narrative constituent structures.

Taken together, these results show that disruptions within a narrative constituent have a greater impact on processing than disruptions between narrative constituents. This suggests that the comprehension of sequential images draws upon a narrative structure that is organized into constituents, analogous to grammatical structure in language.

References

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Cohn, N. (2003). *Early Writings on Visual Language*. Carlsbad, CA: Emaki Productions.

Cohn, N. (In Press). Visual narrative structure. *Cognitive Science*.

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. doi: 10.1016/j.cogpsych.2012.01.003

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

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

Garrett, M. F., & Bever, T. G. (1974). The Perceptual Segmentation of Sentences. In T. G. Bever & W. Weksel (Eds.), *The Structure and Psychology of Language*. The Hague: Mouton and Co.

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8*, 59-85.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324-363.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Holcomb, P., & McPherson, W. B. (1994). Event-Related Brain Potentials Reflect Semantic Priming in an Object Decision Task. *Brain and Cognition, 24*, 259-276.

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.

Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49.

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.

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

Levelt, W. (1970). A scaling approach to the study of syntactic relations. In G. B. Flores d'Arcais & W. Levelt (Eds.), *Advances in psycholinguistics* (pp. 109-121). New York: Elsevier.

Mandler, J. M. (1987). On the psychological reality of story structure. *Discourse Processes, 10*, 1-29.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 111-151.

McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.

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.

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

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.

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

Sitnikova, T., Holcomb, P. J., & Kuperberg, G. (2008). Neurocognitive mechanisms of human comprehension. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding Events: How Humans See, Represent, and Act on Events* (pp. 639-683): Oxford University Press.

Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology, 9*, 77-110.

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

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.