Oxford Handbooks Online

The Neurophysiology of Event Processing in Language and Visual Events

Neil Cohn and Martin Paczynski

The Oxford Handbook of Event Structure
Edited by Robert Truswell

Print Publication Date: Mar 2019 Subject: Linguistics, Semantics, Language and Cognition
Online Publication Date: May 2019 DOI: 10.1093/oxfordhb/9780199685318.013.26

Abstract and Keywords

Events are a fundamental part of all actions that we undertake, discuss, and view. This chapter reviews the growing cognitive neuroscience literature exploring the structure and processing of tacit event knowledge. Overall, this work implies an ongoing process of building up semantic associations relative to their preceding context, which are constrained by active prediction of upcoming content. In turn, reanalysis mechanisms are initiated when the overall structure of the event is violated and/or does not conform to predictions set up by the context. Event cognition thus emerges from a network of conceptual information links with event-specific selectional restrictions and hierarchical organization. Altogether, this literature suggests that the brain uses overlapping mechanisms of comprehension to rapidly process events in both language and visual events.

Keywords: cognitive neuroscience, N400, P600, event segmentation, visual event

22.1 Introduction

EVENTS are a fundamental part of human experience. All actions that we undertake, discuss, and view are embedded within the understanding of events and their structure. With the increasing complexity of neuroimaging over the past several decades, we have been able for the first time to examine how this tacit knowledge is processed and stored in people's minds and brains. Among the techniques used to study the brain, electroencephalography (EEG) offers one of the few ways in which we can directly study information processed by the brain. Unlike functional imaging, whether PET or fMRI, which rely on metabolic consequences of neural activity, the EEG signal is generated by post-synaptic potentials in pyramidal cells which make up approximately 80% of neurons within the cerebral cortex. As such, EEG offers a temporal resolution measured in milliseconds, rather than seconds, making it well suited for exploring the rapid nature of language processing. Though there are numerous ways in which the EEG signal can be analysed, in the current chapter we will focus our attention on the most common measure: event-related potentials (ERPs), the portion of the EEG signal time-locked to an event of interest, such as a word, image, or the start of a video clip.

Each ERP component is characterized by polarity (positive or negative), temporal distribution (both peak and duration), as well as scalp distribution. Though numerous ERP components have been associated with language processing, in the current chapter (p. 625) we will focus primarily on the 'N400', a negative-going deflection lasting two to three hundred milliseconds, peaking around 400ms post-stimulus onset, and the 'P600', a positive-going deflection starting between 400 and 600ms post-stimulus onset and lasting several hundred milliseconds. As we will discuss, though these ERP components had initially been associated with language processing, recent research has suggested that similar brain responses are also involved in the processing of visual events. Such findings have implications both for the understanding of event structure more broadly, and for the domain generality of processing in the human brain.

22.1 Events and basic sentence processing

The ERP component first discovered as relevant for language processing is the 'N400' effect. Though the N400 was first reported by Kutas and Hillyard (1980) in response to semantically anomalous sentence-final words (*He spread the warm bread with *socks*), modulation of the N400 occurs even in the absence of violations of semantic content. One of the first factors found to be strongly associated with size of the N400 is how predictable a word was within a given context (Kutas and Hillyard 1984). The more unexpected a word is within a given context, the larger the N400 amplitude, regardless of the degree of constraint for a sentence's frame (Kutas and Hillyard 1984). Importantly, the plausibility of a word itself does not appear to modulate the N400 (Kutas and Federmeier 2011). Thus, a highly plausible word can nonetheless elicit a larger N400 amplitude when it is unpredictable, versus predictable, within a given context.

Predictability in the sense meant here need not mean the prediction of a specific lexical item, nor the 'preactivation' of information prior to reaching a given lexical item (DeLong et al. 2014, van Petten and Luka 2012). Rather, modulation of the N400 suggests that what is 'predicted' are semantic relationships. Priming studies have found that the N400 is modulated by semantic associations (e.g. dog–CAT; Deacon et al. 2004), shared category membership (e.g. tulip–ROSE; Grose-Fifer and Deacon 2004), as well as semantic feature overlap (e.g. wig–MOP), all of which attenuate the amplitude of the N400. This modulation can be even stronger within a sentence or discourse context. Federmeier and Kutas (1999*a*,*b*) have repeatedly found that words which share features with the most expected word within a given discourse frame evoke an attenuated N400 compared to words with little feature overlap (e.g. *They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of palms / *pines / *tulips*).

Although we have discussed modulation of the N400 in terms of predictability, the idea of prediction in language remains controversial (DeLong et al. 2014, van Petten and Luka 2012).[1] (p. 626) Specifically, many authors have argued that the attenuation of the N400 by semantic relatedness is due to spread of activation within the semantic network (Anderson 2013). In this account a word, or larger discourse context, activates its 'neighbours,' thus facilitating their semantic integration. In a recent study, Lau et al. (2013) sought evidence that prediction might offer additional explanatory power above and beyond the passive spread of activation by using a priming study in which the authors manipulated a proportion of related word pairs within testing blocks. While passive spread of activation would be expected to be independent of block type, prediction would be expected to be more likely in blocks with a high proportion of semantically related word pairs. The authors found the classic attenuation of the N400 to words primed by related words, compared to unrelated words, only in high-related proportion blocks. Indeed, in low-proportion blocks, semantic association failed to modulate the N400 in any appreciable way. This study thus strongly suggests that the N400 is modulated by active prediction of semantic relationships, rather than simple passive spread of activation.

This modulation of the N400 by the degree to which participants engage in prediction thus offers a means of investigating the types of information that are used in constructing event representations. For example, Ferretti et al. (2007) found that, while the typicality of a location for an event described by a verb in the imperfective aspect modulated the amplitude of the N400, this was not the case following verbs in the perfective aspect. This finding aligns with the idea that the imperfective aspect focuses on the context related to the ongoing development of the event, and thus locations, paths, or tools, etc. (cf. Morrow 1990), while the perfective aspect focuses on outcomes of the event. More recently, Becker et al. (2013) found that while the N400 was attenuated to words related to a previously described accomplishment (e.g., packing a lunch) when it was in the imperfective aspect, versus perfective aspect, no such attenuation was observed for activities (e.g., exercising). This finding is again in line with the idea that the imperfective aspect focuses attention on intermediate stages of events. Specifically, for accomplishments the intermediate stage is distinct from the end stage, due to their inherent endpoint, while this distinction is absent for activities, which are more temporally uniform. Interestingly, this modulation of the N400 was only observed in scenarios which described an event of short duration intervening between the critical verb phrase and probe words. When the intervening event was longer, grammatical aspect failed to modulate the N400 evoked by probe words following both activities *and* accomplishments. As the intervening event made no explicit reference to time, these findings suggest that real-world knowledge is used to construct event representations, including their temporal structure.

(p. 627) Indeed, numerous studies have demonstrated that real-world knowledge plays a critical role during online build-up of complex event representations. For example, Chwilla and Kolk (2005) found that the N400 to the third word of a triplet was attenuated when this triplet formed a coherent 'script,' such as DIRECTOR–BRIBE–DISMISSAL, compared to the same word appearing at the end of a nonscript triplet, such as VACATION–TRIAL–DISMISSAL. Because the semantic association between the final word and the preceding two words was constant between the two conditions, the finding suggests that not only is real-world knowledge accessed rapidly during online comprehension, this can occur even with minimal linguistic input. As might be expected, the effect of real-world knowledge on the build-up of complex event representations is even more profound within a more complete discourse context. This was demonstrated by Sanford et al. (2011), who found that anomalous words which were nonetheless consistent with the overall schema described in short discourse (e.g. *Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year sentence was given to the* ***victim****, but this was subsequently appealed*) evoked an N400 which was indistinguishable from that of the contextually appropriate word (e.g. *accused*). Interestingly, when participants detected such anomalies, the implausible words evoked a robust P600, which we shall discuss next.

Elements of events based on real-world knowledge can also be activated beyond the words expressed in a discourse. Words related to a discourse frame (e.g. *jacket*) evoke a more attenuated amplitude N400 than words like *towel* when mentioned following a discourse context about winter (Metusalem et al. 2012), even when both appear as incongruous to the sentence (e.g. *A huge blizzard ripped through town last night. My kids ended up getting the day off from school. They spent the whole day outside building a big snowman / *jacket / *towel in the front yard.*) Thus, a network of associations related to event knowledge may become activated even if they are not expressly mentioned in a discourse.

The P600 was first described by Osterhout and Holcomb (1992; see also Hagoort et al. 1993) in response to both syntactic ambiguities and outright syntactic violations. Specifically, a P600 was observed to the word *to* in sentences such as *The broker persuaded to sell the stock* regardless of whether or not the sentence could resolve to be syntactically well-formed (… *was sent to jail*). Interestingly, the magnitude of the P600 evoked in such garden-path sentences was found to be dependent on the subcategorization biases of the preceding verb (Osterhout et al. 1994), being progressively smaller as intransitive bias decreased.

Although the P600 was originally found to be elicited by a variety of syntactic factors, including agreement violations, subcategorization violations, and phrase structure violations (Hagoort et al. 1993), nonetheless it soon became obvious that the P600 was dependent on semantic content of sentences as well. For example, Gunter et al. (2000) found that the P600 evoked by syntactic violations occurring on words which were highly expected within a given context was greater than when the violations occurred on less expected words. Indeed, some studies have found that morphosyntactic violations occurring in so-called Jabberwocky sentences, in which all content words are replaced (p. 628) by pseudo-words, failed to evoke a P600 effect (but see Ericsson et al. 2008 and Hahne and Jescheniak 2001).

Indeed, the P600 has also been elicited in situations without outright violations of syntax. Kuperberg et al. (2003) found that animacy violations, such as *For breakfast the eggs would only *eat…*, evoked a P600 which was similar to that evoked by outright morphosyntactic violations. Since then, similar findings have been reported by other groups (e.g., Hoeks et al. 2004, Kim and Osterhout 2005) not only for animacy violations occurring on verbs but noun phrases as well (Paczynski and Kuperberg 2011, 2012). However, outright violations of verb animacy constraints are not necessary to evoke a P600. As already mentioned above, Sanford et al. (2011) reported a P600 to semantic violations which violated schema-based knowledge. Similarly, Kolk et al. (2003) likewise found a P600 to violations of schema-based knowledge (e.g., *De kat die voor de muizen vluchtte rende door de kamer* ['The cat that from the mice ?fled ran across the room']). Kemmerer et al. (2007) reported a P600 to nonstandard ordering of adjectives, which were nonetheless syntactically licensed (e.g., *brown big dog*). Taken together, these findings suggest that the P600 is elicited in response to structural violations within linguistic input, whether the 'structure' is syntactic or conceptual. This may explain why P600 effects have been observed to violations occurring outside standard linguistic input, as we shall discuss in the subsequent sections.

Before continuing on to nonverbal event processing, it is important to note that the N400 and P600 are not the only ERPs which have been associated with event processing. For example, several studies have reported a sustained, late anterior negativity in response to complex event processing. Bott (2010) and Paczynski et al. (2014) reported a sustained late negativity to additive and iterative aspectual coercion, respectively. Baggio et al. (2008) reported a similar negativity to events (e.g. spilling coffee on paper) which prevented an achievement reaching its natural conclusion (e.g. writing a letter). Finally, Wittenberg et al. (2014) reported a sustained, late negativity to light verb constructions, such as *Julius gave Anne a kiss*. The common theme amongst all these studies is that the meaning of the overall event being described requires a complex mapping between syntactic and semantic information. The authors of these studies have suggested that this sustained negativity may reflect an increase in working memory demands necessary to generate an appropriate overall event representation. It is unclear whether analogous results have appeared in the visual domain as of yet, and we will not discuss this anterior negativity further, but rather simply mention it for the sake of completeness.

22.3 Nonverbal event processing

Research over the past decade has shown that event processing in the visual, spatial domain appears to engage similar neurocognitive processing as in the verbal, linguistic (p. 629) domain. This work has studied event comprehension using movie clips of real-world events or visual narratives in the form of sequential images (like comic strips) or filmed movies. Below, we review this growing literature with an emphasis on its relations to linguistic processing.

22.3.1 Meaning and structure (N400 and P600s)

As in Kutas and Hillyard's (1980) classic experiment introducing anomalous words at the ends of sentences, the first study looking at the neurocognition of sequential narrative images found a larger N400 to incongruous final images of a visual sequence than to congruous endings (West and Holcomb 2002). In subsequent work, this effect has been found consistently in experiments where images violate the expected meaning of the narrative or events in sequential images or movie clips (see Amoruso et al. 2013 for review). As in language processing though, N400 effects appear in contexts beyond violations. Larger N400s have also been reported to congruous, yet unpredictable, visual events when compared with more predictable events (Giglio et al. 2013, Reid et al. 2009, Reid and Striano 2008). In addition, just as the N400 is attenuated across the ordinal position of a sentence (van Petten and Kutas 1991), the N400 decreases in amplitude across the ordinal position of sequences of coherent visual events (Cohn et al. 2012, Giglio et al. 2013), yet no such attenuation occurs to sequences lacking narrative and/or semantic coherency (Cohn et al. 2012).

These studies suggest that, as in studies of sentences, the processing of visual events is modulated by the predictability of an event with regard to its prior context, and that coherent visual sequences allow for the build-up of meaning across the course of the sequence. These findings are consistent with the idea that the N400 reflects the access of semantic meaning in memory in reference to an incoming stimulus's prior context (Kutas and Federmeier 2011), though the cortical networks involved in such memory may differ between language and visual event comprehension (Balconi and Vitaloni 2014, Ganis et al. 1996, West 1998).

Also similar to language, violations of structure in visual events evoke a positivity similar to the P600, often preceded by an N400 (see Amoruso et al. 2013 for review). Following the discoveries that the P600 in sentences was not strictly sensitive to syntax (e.g., Kuperberg et al. 2003), researchers set out to discover if comparable manipulations could yield P600 effects in visual events that logically followed the same violations that evoked positivities in language. In the first of such studies, Sitnikova et al. (2008*b*) presented participants with silent movie clips of everyday events (like cutting bread) where the final clip either showed a congruous ending (the bread being cut), a semantically anomalous ending (someone ironing), or the expected event being carried out with an incongruous object (the bread being cut with an iron), as in Figure 22.1. Both violations evoked N400 effects, but a late posterior positivity consistent with the P600 also appeared to the events which involved unexpected objects. These findings lead to (p. 630) the idea that this P600 effect indexed a process of reanalysis needed to integrate the event with the selectional restrictions of its objects, similar to the selectional restrictions in argument structure of a verb in sentence processing (Kuperberg 2007, Sitnikova et al. 2008*a*,*b*).

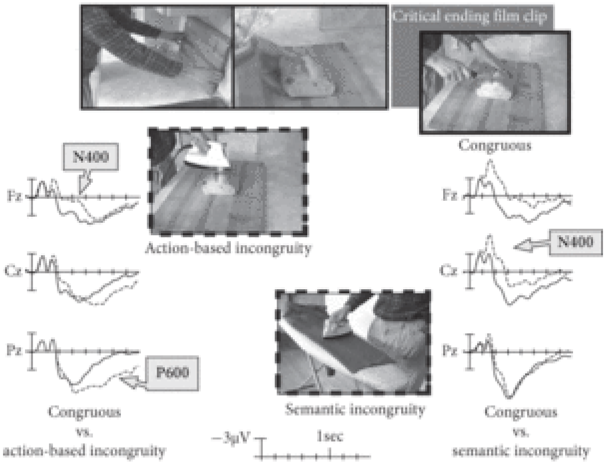

Figure 22.1 Event-related potentials to semantic incongruities and action-based incongruities in a visual event sequence.

Adapted from Sitnikova et al. (2008*b*)

Subsequent research has reinforced that P600s appear in the processing of manipulated visual events. Late posterior positivities appear when participants view actions where the events are not carried out normally (de Bruijn et al. 2007, Giglio et al. 2013), where objects and/or hand position mismatch their corresponding events (Bach et al. 2009, van Elk et al. 2012), or when the motion associated with an event is graphically represented anomalously (Cohn and Maher 2015). P600 effects also do not seem to be limited to monomodal aspects of an event’s structure. They have also been found when participants view visual events where auditory sounds mismatch with visually depicted events, such as video of fireworks exploding co-occurring with the sound of a droplet hitting water (Liu et al. 2009, 2011).

It is worth noting that visual event representations have been shown to influence online language processing. Vissers et al. (2008) found that in sentences which mismatched the spatial relationship between two previously presented objects (e.g., ■▲ followed by *De driehoek staat voor het vierkant* [‘The triangle stands in front of the square’]), a P600 was evoked by the incongruous relational prepositional phrase (e.g. *in front* vs. *behind*). More importantly, recent work has suggested that visual representation of events can rapidly influence the online interpretation of spoken language. (p. 631) Knoeferle et al. (2008) utilized the relatively free ordering of verb arguments in German transitive constructions to examine the timing of syntactic disambiguation in noncanonical (object-first) constructions in which the initial argument was syntactically ambiguous. When sentences were presented on their own, the authors reported a classic P600 on the second noun phrase when its case-marking disambiguated the sentence to the noncanonical OVS construction versus those which disambiguated to a canonical (SVO) construction. However, when participants simultaneously viewed pictorial event representations in which a central figure was an Agent of one action (e.g., painting) and a Patient of another distinct action (e.g., washing), a P600 was observed in ambiguous noncanonical sentences at the verb, rather than the second noun phrase. These findings suggest that people readily utilize both linguistic and extra-linguistic information to rapidly generate overall event representations.

The appearance of P600 effects in contexts outside of language have motivated theories that they do not merely reflect processing of linguistic structure. Recent work has hypothesized that the P600 effect may index prolonged attempts to integrate an input into a structure in the face of a ‘prediction error’, which causes a process of updating or revising the wider representation of context (Christiansen et al. 2011, Kuperberg 2013). Such ‘prediction error’ is caused by the mismatch between top-down predictions generated from a broader context and the bottom-up content of an incoming stimulus. In the case of visual event representations like cutting bread with an iron, the selectional restrictions of the event provide constraints on the expected way in which it should be carried out (particular objects acting in thematic roles, context, etc.), and violation of these constraints demands a reanalysis (Kuperberg 2007, Sitnikova et al. 2008*a*,*b*). This same reanalysis or integration may take place for the selectional restrictions of events in language, most overtly seen in the violation of argument structure related to verbs. Additional support for this domain-general perspective comes from the appearance of P600s to content violations in visual scenes which did not represent events (Võ and Wolfe 2013), violations of the structure of sequential images (Cohn et al. 2014, Cohn and Maher 2015), and violations to the ‘grammar’ of music (Koelsch and Siebel 2005, Patel 2003, Patel et al. 1998). Such findings reinforce the idea that the P600 is not tied to linguistic processing per se, being evoked across domains in response to violations to structure that may be highly constrained by contextual expectations.

22.3.2 Hierarchy in event structure

This emphasis on prediction error aligns nicely with experiments looking at the neurocognition of event segmentation. Research has consistently shown that both adults and children distinguish between violations before or after the boundaries of visual events (Baldwin et al. 2001, Pace et al. 2013). The most substantial research on event segmentation and the brain has been carried out by Jeffrey Zacks and colleagues. Over several experiments, Zacks and colleagues have found that participants consistently (p. 632) agree on where to segment events at both coarse- and fine-grained levels (Zacks et al. 2001*a*,*b*, 2006, 2011). Because the boundaries of fine-grained segmentation both align with and fall within the coarse-grained segments, these results have suggested that events are organized hierarchically. In studies using fMRI, brain activation from passive viewing of visual events correlated temporally with the boundaries between events later identified through a segmentation task (Zacks et al. 2001*a*). The strongest brain activation for event boundaries came in bilateral posterior occipital, temporal, and parietal cortices, and the right lateral frontal cortex. This activation in frontal areas aligned with previous findings of impairment in coarse-grained event segmentation by patients with damage to the prefrontal lobe (Zalla et al. 2003) and with proposals for dedicated event processing in the prefrontal cortex (Wood and Grafman 2003). Additionally, the strongest activation from posterior regions indicated activation in the brain regions associated with the processing of motion, which appear to be sensitive to event boundaries and motion, even when they are carried out by abstract geometric shapes (Zacks et al. 2006).

Zacks and colleagues argue that the segmentation of events relies on the prediction error generated by discontinuities in the perceptual stream of information, such as changes in a figure's posture. Greater discontinuities, such as changes in motion or location, yield the interpretation of more significant event boundaries (Zacks and Swallow 2007). Yet, discontinuities alone cannot fully account for the comprehension of event segments, since brain activation associated with segmentation of visual events has been shown to begin substantially *prior* to viewing an actual event boundary (Zacks et al. 2001*a*). This pre-activation has been taken as evidence for anticipation of the upcoming event boundary, and further research has found that faster and more accurate conscious predictions of a subsequent event occur when predictions are made while viewers remain within an event segment than when the boundary between events has been crossed (Zacks et al. 2001*a*). Zacks and colleagues argue that such predictions allow upcoming events to be responded to faster and with fewer processing costs (Kurby and Zacks 2008).

Nevertheless, the question still remains as to what allows such predictions to occur, since prediction error due to perceptual discontinuities can only happen *after* a boundary has already been reached. Such predictions may be facilitated by knowledge structures about events that divide segments into preparatory actions followed by a completed event (Jackendoff 2007, Lasher 1981). Some recent behavioural work has examined this idea by contrasting Agents (doers of actions) in preparatory actions with Patients (receivers of actions) in more passive states. When participants viewed a sequence of static images depicting events at their own pace, increased viewing times appear to images of Agents depicting preparatory actions than Patients in passive states (Cohn and Paczynski 2013). These costs may be attributable to the predictions made by the postural cues of these preparatory actions (Aglioti et al. 2008, Kourtzi and Kanwisher 2000, Springer et al. 2013), and indeed, when viewing these images in isolation, participants agreed upon more predictions for subsequent actions for preparatory Agents than the Patients. Furthermore, the 'frontloading' of information (p. 633) made by preparatory actions appears to facilitate subsequent event processing, since viewing times to completed actions were faster when following a preparatory Agent than a passive Patient. Together, these findings suggested that Agents aid in the building of event structures. This preference for building event structures from Agents may also offer an interpretation for why languages prefer Agents prior to Patients in canonical and novel sentence structures (Goldin-Meadow et al. 2008, Kemmerer 2012).

These types of anticipatory effects for events may also be modulated by expertise (Aglioti et al. 2008, Amoruso et al. 2014). For example, Amoruso et al. (2014) showed videos of normal tango dancing routines or those with errors to expert and beginning tango dancers along with naïve controls. Prior to the actual errors in the dance routines, deviations appeared in ERP amplitudes for routines with errors in contrast to dance routines without errors. However, these large differences were shown only for expert dancers, while intermediate differences in amplitudes appeared for beginning dancers, and no differences appeared for novices. That is, experienced dancers were anticipating these errors prior to their full manifestation, and this effect was modulated by expertise with these actions. Such work is consistent with other findings of more widespread brain activation for anticipation of actions by expert than naïve viewers (Abreu et al. 2012, Amoruso et al. 2014).

Finally, some evidence has suggested that hierarchic structure can frame the presentation of visual events with a narrative structure. Cohn (2013*b*) has proposed that visual narratives depicting sequential visual events (such as those found in comics) are organized with a hierarchic ‘narrative grammar’ that is separate from semantics and organized into hierarchic constituents. In initial research examining this structure, we compared sequences with a felicitous narrative structure but no semantic meaning between images (visual narrative analogues to sentences like *Colourless green ideas sleep furiously*) with normal sequences, scrambled sequences of random images, and sequences that maintained semantic associations across images (like baseball) but had no coherent narrative structure (Figure 22.2a). First, in a behavioural experiment, shorter reaction times appeared to target panels within normal sequences—which had both narrative structure and semantic associations between panels—than to targets in scrambled sequences, which had no coherent narrative or meaning (Figure 22.2b). However, intermediate times appeared to target panels in sequences that had only narrative or semantic associations, suggesting that these structures gave an advantage to processing, but not as much as their combination.

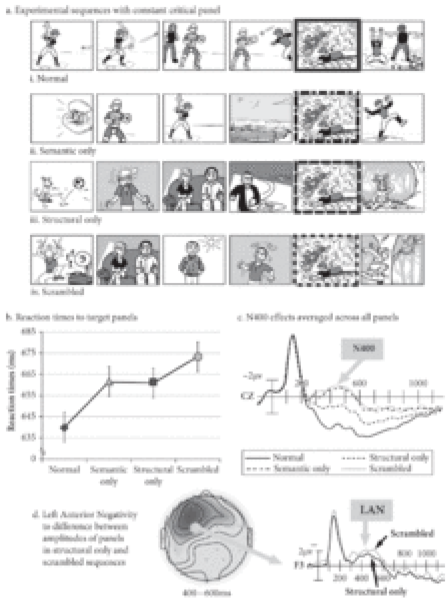

Figure 22.2 Experiments looking at the contributions of narrative structure and semantic associations in the processing of sequential images.

From Cohn (2013*a*: 123, based on Cohn 2012)

A second experiment then analysed these same sequences using ERPs. Larger N400 effects appeared to images in scrambled sequences and those with only narrative structure than to those in normal sequences or sequences with semantic associations alone (Figure 22.2c). However, no difference appeared in the amplitude of the N400 effect for images in fully scrambled sequences and those with only narrative structure, despite the fact that a behavioural advantage appeared to these images (Cohn et al. 2012). Rather, images in these structure-only sequences differed from scrambled sequences with a negativity localized to a left anterior region of the scalp (Figure 22.2d). This distribution was distinctly different from the more widespread negativity of the N400, leading to a (p. 634) (p. 635) tentative proposal that this was a ‘left anterior negativity’ (LAN) effect, typically shown to violations of constituent structure in sentence processing (Friederici et al. 1993, Neville et al. 1991), or on the right side of the scalp ('right anterior negativity', RAN) to violations of musical grammar (Koelsch and Siebel 2005, Patel 2003, Patel et al. 1998).

In follow up work, Cohn et al. (2014) sought to directly examine the constituent structure of visual narratives. This experiment followed the design of classic studies from psycholinguistics which found that the recall and comprehension of sentences with 'clicks' of white noise within grammatical constituents was worse than for when clicks fell at the natural breaks between constituents (Fodor and Bever 1965). Analogously, blank white 'disruption panels' were introduced into sequences that fell either within the narrative constituents of a visual sequence or between constituents. Disruptions within constituents evoked a larger left-anterior negativity than blanks that appeared at the breaks between constituents. It is important to note that disruptions within the first constituent showed a larger amplitude effect than those between constituents, despite the fact that participants had no confirmation that the end of the constituent had been reached, since they had not yet viewed the first panel of the next constituent. This means that participants could not have relied on the linear semantic discontinuities between images to signal a constituent break (e.g., Zacks and Swallow 2007), but rather were making predictions about subsequent structure prior to that break occurring.

Leftward anterior negativities have also been shown directly to depictions of visual actions. In one study, expert basketball players showed a larger negativity to violations of basketball actions than nonviolated actions, an effect that did not appear to nonexperts (Proverbio et al. 2012). Though this effect was reported as an N400, it localized over left anterior electrode sites and showed origins in the left hemisphere through source localization. This distribution more resembles a LAN than the typically widespread anterior distribution of N400s in images that extends across both hemispheres. Thus, it may be possible that such an effect does not reflect the access of semantic networks, as suggested by an N400, but a left anterior negativity sensitive to the structural architecture of event information.

As with the N400 and P600, the appearance of lateralized anterior negativities in the processing of visual narrative, visual events, and music suggests that these effects may not index strictly linguistic processing. In all cases, these effects appear to violations of the structural makeup of these domains, yet unlike the P600, these violations do not appear in the presence of attempts to access meaningful information. Rather, these violations are associated with structure alone, suggesting a preliminary interpretation that the brain may process hierarchic structure across domains in a similar way, indexed by these anterior negativities. Such a view is supported by the idea that anterior negativities in language may originate in Broca's region (left BA44/45) (Friederici 2002, Hagoort 2003), a brain area long associated with syntactic processing, while the RAN shown to violations of musical syntax has also been tied to activation in Broca's region (Maess et al. 2001). Some crossmodal research has also suggested similar underlying cognitive processing for co-occurring violations to both language and music, manifesting as both (p. 636) anterior negativities and posterior positivities (Koelsch et al. 2005).

Concomitant to this view is research showing impaired ability to reconstruct coherent visual narratives and visual events by Broca's aphasics (Bihrle et al. 1986, Fazio et al. 2009) and by healthy adults with BA44 disrupted through the use of transcranial magnetic stimulation (Clerget et al. 2009). Such findings suggest initial converging evidence that similar neural mechanisms processes structure across domains. However, future work directly comparing the processing of language, visual events, and/or visual narrative within the same experimental battery would be necessary for analysing the neurophysiological overlap between these diverse domains.

22.4 Summary

In this chapter we have described numerous ways in which events are comprehended in language and visual events, pointing towards overlapping mechanisms of comprehension between them. Our overall vision for event comprehension may be visualized as in Figure 22.3, where a network of conceptual information links with event-specific selectional restrictions and hierarchic organization. First, comprehension of events, whether presented as linguistic input (Chwilla and Kolk 2005, Sanford et al. 2011), a series of panels in visual narrative (Cohn et al. 2012), or as movies (Sitnikova et al. 2008*b*), rely on a build-up of semantic associations created by the context which precedes them. Evidence for this comes from modulation of the N400, which is attenuated when the event is semantically associated to its preceding context. Second, this process appears to (p. 637) be based on active prediction of upcoming content, again, whether this occurs within a traditional linguistic context (Lau et al. 2013) or in the form of a visual narrative (Cohn et al. 2014, Cohn and Paczynski 2013, Cohn et al. 2012). Finally, when the overall structure of the event is violated and/or does not conform to predictions set up by the context, a P600 is observed regardless of whether the input is linguistic (Kuperberg et al. 2003) or involves a visual depiction of the event (Sitnikova et al. 2008*b*). Taken together, these findings suggest that the brain uses similar neural substrates to rapidly process events independently of the form in which it is presented to the comprehender.

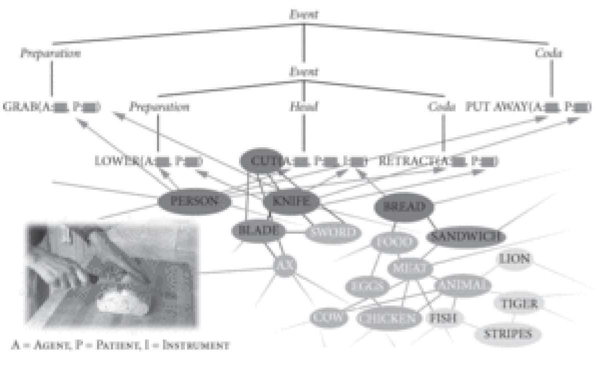

Figure 22.3 Depiction of the network and structure involved with event comprehension.

Acknowledgements

Thanks go to Eva Wittenberg for her support and encouragement in the writing of this chapter.

Since the writing of this chapter, it is with great sorrow that I acknowledge that Dr Martin Paczynski has passed away. He was a scholar with a keen mind for experimentation, a wide range of interests, and insight especially regarding language and event structure. I am happy that this piece, among his other writings, can carry on his memory and scholarly contributions.

Notes:

([1]) Note that 'prediction' in language processing can also extend beyond the comprehension of events. For example, several studies have now shown evidence of 'pre-activation' in syntactic structure when, for example, preceding articles disagree with the predictions of a highly expected word (DeLong *et al.* 2005, van Berkum *et al.* 2005, Wicha *et al.* 2003). For more extensive reviews of 'prediction' and the difference between 'prediction' and 'predictability' in language processing, see DeLong *et al.* (2014) and van Petten and Luka (2012).

Neil Cohn

Neil Cohn is an assistant professor at the Tilburg center for Cognition and Communication at Tilburg University. He is internationally recognized for his research on the overlap of the structure and cognition of sequential images and language. His books, The Visual Language of Comics (Bloomsbury, 2013) and The Visual Narrative Reader (Bloomsbury, 2016), introduce a broad framework for studying visual narratives in the linguistic and cognitive sciences. His work is online at www.visuallanguagelab.com.

Martin Paczynski

Martin Paczynski, until his untimely death in 2018, was a Cognitive Neuroscientist at Wright State Research Institute, affiliated with Wright State University. He received his PhD in Psychology from Tufts University in 2012, focusing on ERP studies of event structure, aspect, and animacy. He subsequently worked on the effects of low-intensity stress (and its amelioration) on perceptual and cognitive performance. Memorials can be found at http://paczynski.org.