Brain & Language 120 (2012) 135–162

Contents lists available at SciVerse ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

On the early left-anterior negativity (ELAN) in syntax studies

Karsten Steinhauer [a,b,*], John E. Drury [c]

[a] Centre for Research on Language, Mind and Brain, McGill University, Canada
[b] School of Communication Sciences and Disorders, McGill University, Canada
[c] Department of Linguistics, Stony Brook University, United States

ARTICLE INFO

Article history:
Available online 14 September 2011

Keywords:
Event-related potentials (ERPs)
Syntax
Early left-anterior negativity (ELAN)
Phrase structure violations
P600
N400
LAN
Blocking effects
Syntax-first models
Baseline problems

ABSTRACT

Within the framework of Friederici's (2002) neurocognitive model of sentence processing, the early left anterior negativity (ELAN) in event-related potentials (ERPs) has been claimed to be a brain marker of syntactic first-pass parsing. As ELAN components seem to be exclusively elicited by word category violations (phrase structure violations), they have been taken as strong empirical support for syntax-first models of sentence processing and have gained considerable impact on psycholinguistic theory in a variety of domains. The present article reviews relevant ELAN studies and raises a number of serious issues concerning the reliability and validity of the findings. We also discuss how baseline problems and contextual factors can contribute to early ERP effects in studies examining word category violations. We conclude that – despite the apparent wealth of ELAN data – the functional significance of these findings remains largely unclear. The present paper does not claim to have falsified the existence of ELANs or syntax-related early frontal negativities. However, by separating facts from myths, the paper attempts to make a constructive contribution to how future ERP research in the area of syntax processing may better advance our understanding of online sentence comprehension.

1. Introduction

The focus of the present inquiry is on an influential neurocognitive model of sentence processing, proposed by Friederici (1995, 2002; Friederici & Kotz 2003; Friederici & Weissenborn 2007), and in particular on this model's interpretation of a brain response revealed in studies of syntactic processing using event-related potentials (ERPs), namely the early left anterior negativity or "ELAN". ERP data have strongly contributed to the development and refinement of this serial, syntax-first model which offers a detailed characterization of incremental processing of sentences in terms of three phases which apply consecutively for each individual word. Further, each phase is claimed to be reflected by distinct ERP components within specific latency ranges after relevant types of word information become available.

In **phase-1** (100–300 ms), initial phrase structure (**PS**) building based exclusively on syntactic word category information takes place. Disruptions of this fast, highly automatic, first-pass parse due to word category violations are claimed to be reflected by ELAN effects, which are argued to uniquely index the action of brain systems underlying PS generation (e.g., Hahne & Friederici, 1999).

During **phase-2** (300–500 ms), both (i) morpho-syntactic processing (including feature checking) as well as (ii) lexical/conceptual-semantic integration take place. Processing difficulties in morpho-syntax tend to elicit left anterior negativities (LANs; e.g., verb inflection violations, see Gunter, Stowe, & Mulder (1997) among others) or N400s (e.g., argument structure violations; Friederici & Frisch, 2000), whereas lexical-semantic difficulties generally yield N400 components (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008). These two processing streams take place in parallel, such that LAN and N400 can co-occur if both streams encounter difficulties. However, as both streams need to be licensed by the intact phrase structure generated in phase I, N400s and LANs are claimed to be BLOCKED in presence of a word category violation if the latter co-occurs with semantic or morpho-syntactic violations. In such 'double' violations, ELANs are predicted not to co-occur with LANs or N400s.[1]

In **phase-3** (500–1000 ms or later), different streams of information are integrated. If this integration process encounters anomalies involving the sentence structure, additional controlled (i.e., less automatic) reanalysis and repair processes are required and elicit P600 components (Osterhout & Holcomb, 1992), which have also been referred to as syntactic positive shifts (SPS, Hagoort,

* Corresponding author at: McGill University, Centre for Research on Language, Mind and Brain, School of Communication Sciences & Disorders, 1266 Pine Avenue West (Beatty Hall), Montreal, Quebec, Canada H3G-1A8. Fax: +1 514 398 8123.
E-mail address: karsten.steinhauer@mcgill.ca (K. Steinhauer).

[1] Note that there are cases in the literature of apparent early left anterior negativities for contrasts involving "single" PS-violations (i.e., not "double" violations) which are followed by later negativities, the status of which remains rather unclear (e.g., Neville, Nicol, Barss, Forster, & Garrett, 1991 – see also Footnote 3).

0093-934X/$ - see front matter
doi:10.1016/j.bandl.2011.07.001

Brown, & Groothusen, 1993). Hence, ELANs and LANs are typically followed by a P600, while garden path sentences and complex structures without violations may elicit only P600s.[2]

1.1. Objectives and motivations for the present inquiry

The central contribution of the model compared to others lies in its emphasis on the distinct role of word category information during phase-1 and its unique power to unidirectionally block subsequent processes. Therefore, the link between the elicitation of ELANs and syntactic word category violations as well as the power of PS violations in phase-1 to block the putatively "downstream" semantic (and morpho-syntactic) processes are the most crucial and distinctive features of the model.

Here we aim to critically evaluate two important but under-discussed general problems in the ELAN literature. Despite concerns which have previously been raised elsewhere (see, e.g., Dikker, Rabagliati, & Pylkkanen, 2009; Hagoort, 2003; Hagoort, Wassenaar, & Brown, 2003; Hasting & Kotz, 2008; Lau, Stroud, Plesch, & Phillips, 2006; Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004; Steinhauer & Connolly, 2008; Yamada & Neville, 2007), we are not aware of a single article systematically examining these problems in sufficient detail to demonstrate just how serious they may be.

> Problem #1: Taking existing data in the ELAN literature *at face value*, several of the core claims of the model can be argued to be unsustainable.
> Problem #2: There are serious methodological problems which actually recommend we exercise some caution in taking previous ELAN interpretations in the literature *at face value*.

Our critical discussion is motivated by the strong impact the model has had, and continues to have, on the field. Since its first formulation (Friederici, 1995) the model has received impressive empirical support. It inspired dozens of ERP (and other brain imaging) studies whose additional findings led to more refined recent versions (Friederici, 2002; Friederici & Kotz, 2003; Friederici & Weissenborn, 2007). With some 450 citations to date (ISI Web of Science), the 2002 article alone has been cited once every week on average since it went to press.

This enormous impact stems largely from the strikingly coherent network of empirical findings from a broad range of language-related investigations that all appear to converge on this model, and in particular on the central role of the ELAN. Since the more robust 'syntactic' P600/SPS component *follows* the 'semantic' N400 in time and has been shown (contra initial suggestions, see Hagoort et al., 1993) to not be syntax-specific (see, e.g., Münte, Heinze, Matzke, Wieringa, & Johannes, 1998; Patel, Gibson, Ratner, Besson, & Holcomb, 1998), the ELAN has turned out to constitute *the* electrophysiological bedrock for claims that the syntactic sub-component of language processing constitutes a "module" in Fodor's (1983, 2000) sense (see also Friederici, 1990). Friederici's model has thus played an extremely prominent role in the larger arena of (often otherwise ideological) debates about modularity, information flow, and cognitive architecture. Also of importance is the attractive link that Friederici's model established with Frazier's influential (syntax-first) garden path model (Frazier, 1987).

Further, imaging data have been argued to suggest a direct association between the ELAN component and Broca's area, for some the very brain area for syntax (Friederici, Hahne, & von Cramon, 1998) or even the location where the universal grammar module may implemented (Sakai, Hashimoto, & Homae, 2001). The combined data fit well with those from Broca's aphasics and late second language learners (who do not seem to show an ELAN; Friederici et al., 1998; Hahne, 2001). Finally, the idea of ELAN components reflecting automatic parsing has influenced a number of other neurocognitive models (Clahsen & Felser, 2006; Hagoort, 2003; Ullman, 2001), including Bornkessel-Schlesewsky and Schlesewsky's (2006, 2008, 2009) increasingly influential *extended Argument Dependency Model* (eADM), which differs in many other respects from Friederici's framework.

An obvious strength of Friederici's model, which no doubt also helps to account for its pervasive influence, lies in the clarity of its easy to falsify predictions. But, remarkably, what has happened over the past 15 years is that it has continued to gain in influence *without ever changing any of its central assumptions*. It is difficult to avoid the impression that, even if some of the model's specific claims may not survive the further tests of time, most of its core claims must be more-or-less correct.

2. Critique

The critical discussion that follows comes in three parts. In the first two sections (Section 2.1 and 2.2) we confront Problems #1 and #2 in order. However, independent of the ELAN is the idea that phase-1 violations suppress the engagement of phase-2 processing systems ("blocking") – this is addressed in the third part of our critique (Section 2.3). In addition, in the Appendix, we also provide a tabular summary of the findings of the ERP studies most relevant to ELAN and blocking effects, including both reading (Table A1) and auditory (Table A2) experiments.

2.1. Problem #1: empirical concerns

ELAN responses are violation effects thought to be elicited by clashes involving major syntactic category divisions (e.g., nouns versus verbs). How are these effects investigated? The seminal work of Neville et al. (1991), for example, tested (1a/b) in a reading study and reported (among other, later effects) a very early left lateralized relative negativity for (1a) versus (1b) (their "N125").[3]

(1)	a.	*The scientist criticized Max's ****of*** *proof the theorem.*
	b.	*The scientist criticized Max's proof* ***of*** *the theorem.*

Encountering the possessive-marked proper name (*Max's*) leads human parsing mechanisms to strongly predict that the next word will be an open-class/content element (e.g., the head noun, as in *Max's proof* in (1b)). Instead, in the violation condition (1a), a preposition is encountered that seems to violate the local phrase structure (but see Lau et al., 2006). Since this type of design holds the critical target word constant across conditions in which the preceding context is manipulated, we will refer to this as the context manipulation approach.

A similar context manipulation approach has also been employed in most of the German ELAN studies (henceforth referred to as 'the German paradigm'), both in the auditory and the visual modalities (e.g., Friederici, Pfeifer, & Hahne, 1993; Friederici, Steinhauer, & Frisch, 1999; Hahne & Friederici, 1999), most of which re-

[2] The status of P600-type effects has, however, been undergoing much in the way of theoretical reinterpretations (see Friederici, Mecklinger, Spencer, Steinhauer & Donchin (2001), Kuperberg (2007), and Bornkessel-Schlesewky & Schlesewsky (2008, 2009) for discussion).

[3] The other two effects following the N125 in Neville et al. (1991) were: (i) a left lateralized temporal negativity (300–500 ms) and (ii) a subsequent P600. Note that the label "ELAN" post-dates the Neville et al. (1991) study by nearly half a decade, and subsequent work by Neville and colleagues has actually argued against the idea that ELAN effects index the action of an encapsulated syntactic processor (Yamada & Neville, 2007).

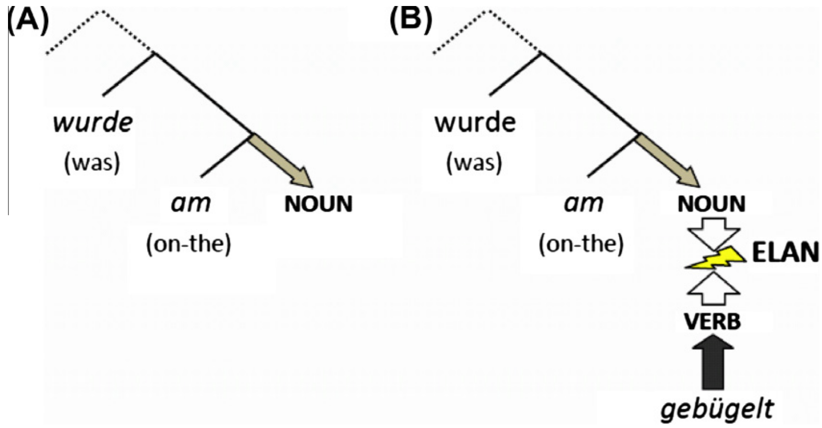

Fig. 1. Illustration of hypothetical pre-(A) and post-(B) target word parsing state in ELAN violation condition (see main text).

ported ELAN components for the target verb (underlined) in example (2a), compared to control conditions such as those in (2b–d):

(2)	a.	*Die Bluse wurde am ****gebügelt***.
		The blouse was on-the ironed.
	b.	*Die Bluse wurde* ***gebügelt***.
		The blouse was ironed.
	c.	*Die Bluse wurde am Montag* ***gebügelt***.
		The blouse was on-the Monday ironed.
	d.	*Die Bluse wurde oft* ***gebügelt***.
		The blouse was often ironed.

The violation in (2a) is created by omitting the required noun within a preposition phrase (PP); the preposition *am*(on-the) is instead followed by a verb form (past participle *ge-bügelt,* where the prefix *'ge'* conveys the critical word category information and is supposed to elicit the ELAN right after word onset). The grammatical control conditions contain no PP at all (2b), an intact PP (2c), or an adverb preceding the participle (2d).

Theoretically, the relevant type of clash in (1a)/(2a) can be conceived as a mismatch between (i) word category (WC) predictions/expectations derived from the current state of the internal representation constructed by human parsing mechanisms on the basis of previous input, and (ii) WC information retrieved from long-term lexical memory corresponding to the actual incoming word that is currently being processed. For example, encountering the preposition *am* in the ill-formed German example in (2a) leads parsing mechanisms to predict the upcoming presence of a noun, as represented in Fig. 1A. However, if the subsequent item (target word) turns out to be a verb (as *gebügelt* in Fig. 1B), the word category information for this word will mismatch with the prediction encoded in the parser's representation of the previous input. It is *this* type of clash which is thought to underlie the ELAN response in phase-1.

2.1.1. Modality independence of ELANs

For ERP researchers and psycholinguists it is important to know if data from reading studies can be generalized to auditory studies and vice versa. On the one hand, such modality independence regarding the ELAN response has not, to our knowledge, ever been defended as a core claim of Friederici's model. Friederici (2002), for example, is explicitly offered as an account of auditory language processing. On the other hand, the approach has nonetheless relied on reference to findings from reading studies (for one example, Neville et al., 1991 is reliably cited in this connection). Discussions of this matter typically arise in the context of why ELANs seem to be less reliably found in reading studies. For example, consider Friederici and Weissenborn's (2007) Footnote 1:

> *Violation type and language being equal, we have argued that the ELAN component reflecting highly automatic syntactic processing of phrase structure building is more likely to be observed during the processing of connected speech than for visually presented language, especially when words are presented individually with pauses of 300 ms or more between each word or when the visual input is hampered by low visual contrast (*Gunter, Friederici, & Hahne, 1999*).*

Despite qualifications of this sort, it seems to be assumed that ELAN effects should be found in both the visual and auditory domains, and there is some evidence of ELAN-like components in reading (Friederici et al., 1999; Neville et al., 1991; Roehm & Haider, 2009, Experiment 1; Yamada & Neville, 2007).

However, *most* reading studies investigating word category violations failed to observe ELANs (see Appendix, Table A1) and either found only later anterior negativities ([L]ANs) after 300 ms (Friederici, Hahne, & Mecklinger, 1996; Friederici & Meyer, 2004; Hagoort, 2003; Hinojosa, Martin-Loeches, Casado, Munoz, & Rubia, 2003; Martin-Loeches, Munoz, Casado, Melcon, & Frenandez-Frias, 2005; Newman, Ullman, Pancheva, Waligura, & Neville, 2007; Roehm & Haider, 2009, Experiment 2), N400s (Federmeier, Segal, Lombrozo, & Kutas, 2000; Gunter & Friederici, 1999), or even relative *positivities* between 300 and 500 ms (Frisch, Hahne, & Friederici, 2004) compared to control conditions.[4] To our knowledge, the absence of ELAN effects in these studies cannot be attributed to factors mentioned in the quote above, such as presentation rate. Further, the implications of the findings of Gunter et al. (1999), where word category violations elicited ELANs only in high visual contrasts (black font on white background) but not low contrasts (dark gray background) remain unclear. This is because most (if not all) other reading studies actually did use high visual contrasts.

A particularly interesting study is that by Gunter and Friederici (1999). They employed the German paradigm (see (2)), but unlike other studies that used a range of German prepositions the authors included only sentences containing the preposition *vom* (by-the). Instead of an ELAN,[5] the authors consistently found N400s across

[4] The observed positivity in Frisch et al. (2004) may be influenced by the 2×2 ANOVA design that included a semantic violation condition as a control; however, their data do not show any indication of an ELAN-like negativity either.

[5] The only (undiscussed) indication of an early negativity (in Experiment 1) is visible already prior to target word onset and must, therefore, be due to spill-over effects from the context (see Section 2.2).

two tasks and argued that, since this particular preposition assigns thematic roles, it "*may have triggered semantic expectation for an agent, rather than syntactic expectation for a particular word category (noun)*" [p. 136]. However, this line of argument seems to be in conflict with core assumptions of Friederici's model. First, the word category violation (preposition followed by a verb) should have prevented theta role assignment in the first place. Second, the presence of an N400 in *absence* of an ELAN suggests that a phase-2 process (thematic role assignment) blocked the 'automatic' syntactic processing in phase-1 – which should be entirely impossible within the framework.

In sum, ELAN effects in reading studies seem to be the exception (see Table A1 in Appendix A). However, it would be entirely reasonable to respond that modality independence is not a crucial claim of the model. For example, it is conceivable that written language is processed in a less automatic manner than speech because the human language faculty evolved for speech only (whereas reading and writing are skills that need to be learned later in life). Moreover, it could be that ELAN-like effects in reading studies are limited to short closed-class/function words (as in Neville et al., 1991; Yamada & Neville, 2007; and see also Dikker et al., 2009),[6] whereas content words due to their complexity or word length may account at least for delays of these effects (see Friederici & Weissenborn, 2007).

2.1.2. ELAN specificity for "outright" violations

This idea has been explicitly offered to explain failure to replicate ELAN effects in terms of a failure to create outright violations on the critical target words (Friederici, 2002; Friederici & Weissenborn, 2007).

> "*There are some reports of experiments using ERPs that do not show ELAN effects in response to syntactic violations. We have demonstrated [...] that most of these studies used sentence material that did not contain outright syntactic violations, rather they contained correct, although unusual, structures*" (Friederici, 2002).

In other words, subsequent to their respective target words, sentences in those studies that failed to elicit ELANs were still compatible with a syntactically permissible continuation, whereas all studies that *did* elicit ELANs involved local PS violations that categorically ruled out any grammatical continuations.

However, this claim is too strong. First of all, the studies cited by Friederici as examples of failing to elicit ELANs for this reason (e.g., Ainsworth-Darnell, Shulman, & Boland, 1998; Hagoort et al., 1993) were *reading* studies which, as just noted above, have not proven to reliably elicit ELANs even when employing the German paradigm. Second, and most importantly, the German violation paradigm itself does *not* meet the outright violation criterion. As illustrated by Steinhauer and Connolly (2008), most sentences used by Friederici and colleagues to create word category violations (see example (2)) can actually be continued such that a perfectly grammatical sentence results (3a):

(3) a. *Die Bluse wurde am* ***gebügelt*** *noch festlicher wirkenden Jackett mit Nadeln befestigt.*
(The blouse was to-the ironed even more festive seeming jacket with pins fixed.)
(*Paraphrase*: The blouse was pinned to the jacket which, after being ironed, appeared even more festive.)

b. *Die Bluse wurde am gebügelt****en*** *Jackett mit Nadeln befestigt.*
The blouse was to-the ironed jacket with needles pinned.

As German past participles can generally (with few exceptions) also be used as modifiers (adjectives and adverbs) that are entirely grammatical after a preposition (as in (3a)), the German paradigm used by Friederici and colleagues turns out to have the exact same problems as those studies that failed to elicit ELANs: the structure may be unusual, but it certainly does not contain an outright word category violation on the target word. One could argue that sentence final punctuation (in reading studies) and sentence final prosodic cues (in auditory studies) may prevent the brain from considering any further continuation (thereby ruling out a permissible grammatical structure), but this fact in itself indicates that word category information per se is not sufficient to signal the violation.[7] Moreover, sentence-final prosody would render *any* single word following the preposition ungrammatical, including the expected noun (*Die Bluse wurde am* **Montag.*/ The blouse was on-the *Monday.).

Another issue, illustrated in example (3b), results from the fact that the *type* of linguistic "violation" in (2a) is not necessarily clear and can very well be interpreted as a morpho-syntactic agreement error (rather than a word category violation), which should elicit later LAN components (i.e., 300–500 ms). That is, adding the inflectional morpheme *'en'* to '*gebügelt*' renders this sentence beginning into a structure (a PP including an inflected adjective) that is not only grammatical but also highly frequent in German.

This ambiguity raises an important question, which to our knowledge has so far received little attention: *How does a reader's (or listener's) brain determine online what kind of violation/anomaly it is actually dealing with?* The crucial point here is that right at the position of the target word (i.e., where the ERP effects are being elicited), no information is yet available about subsequent words. However, what follows further downstream in a sentence may

[6] This general idea is also consistent with data from a recent MEG study (Dikker et al., 2009) where contrasts like those examined in Neville et al. (1991) were compared alongside stimuli whose deviance turned on either the presence of a bound morpheme (e.g., *The discovery was in the *report****ed***) or the absence of a bound morpheme (e.g., *The discovery was ** ***report_***). Interestingly, both the cases modeled on Neville et al. (1991) stimuli and the cases which involved inflection where it was not expected elicited an enhanced M100 component which is viewed to be the magnetic equivalent of the N100 component in ERP studies. Therefore, top-down syntax-based predictions may be understood to generate expectancies at "perceptual" levels (e.g., visual word form representations). Clashes between these predictions and the actual bottom-up visual input may be what these early effects are truly indexing. Thus, for example, this type of effect may correspond to ERP observations such as Neville et al.'s (1991) "N125" effect (i.e., superimposing the N100).

[7] Such effects of 'sentence final' prosody supporting a phrase structure violation were studied by Eckstein and Friederici (2005, see also 2006), however employing other sentence structures. They reported a right-anterior negativity ("RAN") around 400–600 ms post target onset followed by a P600 for ungrammatical sentences such as "*Peter weiss dass der Onkel *Kuchen.*" ("Peter knows that the uncle *cake."). The RAN was found *only* when (i) the target noun (e.g., *Kuchen*) was not followed by a subsequent word (i.e., no RAN in their *contrast pair 1*) and when the target noun in addition also (ii) contained sentence final prosody (as compared to non-final prosody; cf. their *contrast pair 4*). One detail worth mentioning is that the "RAN"-negativity was visible not only over the right but also over the left hemisphere (cf. their Fig. 3b, right panel; and the 450–550 ms voltage map in their Fig. 4). In line with this pattern, the global ANOVA only revealed a frontal but not a right-lateralized negativity (no interaction with factor hemisphere); thus – in contrast to what the label RAN seems to suggest – the effect was actually quite similar to some bilateral syntax effects that have been discussed as ELAN-like components (e.g., Hasting & Kotz, 2008; Rossi, Gugler, Hahne, & Friederici, 2005, 2006). Similarly, the 2006 study by Eckstein and Friederici also found only a broadly distributed negativity for prosodic mismatches which was neither anterior nor right-lateralized, but which nevertheless was discussed as being similar to a RAN. In other words, the strict distinction between *left-anterior* syntactic and *right-anterior* prosodic negativities is largely unsupported by empirical data. In fact, syntactic ELAN effects reported in the 2006 study (in addition to the prosodic effects) were found at left and right *temporal* rather than left-anterior sites.

actually be critical to determine the type of anomaly. Thus, from a linguistic perspective, an ungrammatical sentence such as *"Die Bluse wurde am *gebügelt Jackett mit Nadeln befestigt"* would arguably be categorized as a morpho-syntactic rather than a word category violation. This is so, simply because the easiest way of correcting the mistake is by adding the inflectional morpheme *"en"* to "*gebügelt*", given the remainder of the sentence. However, according to Friederici's model, the human parser would be expected to interpret this as a word category violation and block further processing (even though this would potentially speed up the online correction). A possible way of dealing with this kind of ambiguity within Friederici's model would be to assume that despite the temporary blocking of phase-2 for the target word *"gebügelt"*, the parser may nevertheless initiate (in parallel) phase-1 for the subsequent word (i.e., *"Jackett"*) and use this information to speed up the reanalysis in phase-3 of the (preceding) target word. We are not aware of any study that has investigated these issues.

2.1.3. Strategies, seriality, and the temporal availability of information

Another claim about phase-1 is that it involves the engagement of what have elsewhere been referred to as modular processes, which are domain specific, encapsulated, fast, automatic/mandatory, etc. (Fodor, 1983, 2000). Important evidence to support this perspective comes from Hahne and Friederici (1999). There the relative proportion of PS violations versus correct control sentences was manipulated across two experiments (20:80 in Experiment 1; 80:20 in Experiment 2). The rationale was that in Experiment 2 subjects could develop strategies and anticipate the structural anomaly.[8] As expected for a component reflecting automatic 'bottom-up' processes, the ELAN was not affected by this manipulation. In contrast, the P600 – assumed to reflect 'controlled' revision processes – was basically absent for violations in the 80:20 experiment. This immunity to strategic influences has been argued to be consistent with the idea that ELANs are indeed indexing the activity of a system that meets the criteria of a Fodorian module. However, we find this to be less than clear when considered together with a range of other facts.

First, consider another idea which is difficult to doubt: the onset latency of ELANs should be expected to vary with the input availability of word category information (especially in auditory studies). Clearly we expect a relationship between the temporal availability of relevant information in the input and corresponding onset latencies of ERP effects (for *any* kind of linguistic violation, not just for phrase-structure). So, is this fairly clear expectation borne out by the data? Two studies relevant to this issue are Friederici et al. (1996) and Friederici et al. (2004). Unlike other studies that used prefixes *ge-/be-* to mark verb forms (past participles) and found ELANs within 200 ms after *target word onset*, these auditory studies used participles beginning with prefix "*ver-*", which is also compatible with nouns and therefore does not provide early word category information. Instead, word category information becomes available only after the word-final *suffix* is processed (e.g., verb *veredel-t* versus noun *Veredel-ung*,[9] English: refin-ed *versus* refinement). In both studies the authors reported the expected *delayed* ELAN effects (after 300 ms post word onset) for violation conditions. Intriguingly, once time-locked to the onset of the critical suffix (i.e., to the – *t* in *veredel-t*) the ELAN component was found within a time range of less than 200 ms (in fact, after no less than 50 ms in the 1996 study). These findings seem to confirm the idea that the timing of ELAN responses is dependent on when precisely the word category information becomes available in the speech stream.

However, the authors' account for the differences in ELAN latency between these studies and others using different prefixes (e.g., *ge-/be-*) cannot be maintained. This is for the simple reason that *ge-/be-* also do *not* provide any reliable information that could be used to distinguish verbs from other word categories, including nouns of all three grammatical genders (*der Ge-burtstag* [birthday], *das Ge-bäude*[building], *die Ge-legenheit*[occasion]; *der Be-weis*[-proof], *das Be-gräbnis*[funeral], *die Be-deutung*[meaning]), adjectives (*ge-scheit*[smart]; *be-liebt*[popular]), and adverbs (*ge-nau*[exactly]; *be-trächtlich*[considerably]), all of which are perfectly grammatical after a preposition (see also example (3) above).[10] As prefixes '*ge-/be-*' cannot signal the word category violation, ELANs at word onset must reflect something else.

But what does this have to do with the point raised above regarding the alleged immunity of the ELAN to strategic influences? The connection is this: the main difference between the experiments rather seems to have to do with whether other (grammatical) word categories starting with the critical prefixes were *actually included in the experiment* (Friederici et al., 1996, Friederici, Gunter, Hahne, & Mauth, 2004) or not (e.g., Hahne & Friederici, 1999, 2002). From this perspective, the data suggest that whether a prefix was indicative of the word category simply depended on the specific experimental context and therefore – importantly – on either some kind of (top-down) strategic processing (in the sense of Hahne & Friederici, 1999) or implicit learning. In studies that did not include nouns starting with *ge-/be-*, it is in fact conceivable that over the course of the experiment participants would learn the association between these prefixes and a specific word category, i.e., verbs.[11] As a consequence, it may be that these prefixes were indeed predictive of word category, but only as a function of their particular distribution in the experiment (i.e., in absence of *ge-/be-* noun forms; see footnote 9).

But, if the foregoing is correct, *this is just to say that the claim that ELAN effects are independent of strategic influences* (or implicit/statistical learning) *cannot be true*. Further, this may also point towards an important qualification of the model's applicability to *normal* language processing (i.e., its ecological validity outside of a laboratory).

Moreover, there are indications in these studies of ELANs time-locked to the word *onset* (i.e., prior to the critical suffix (−**t**) that signaled the WC violation). Thus, Friederici et al. (1996) reported an ELAN at 370 ms relative to word onset, but the ERPs actually show a sustained negativity with an onset which appears to be as early as 200 ms (that is, just after the somewhat unusual post-stimulus 0–200 ms baseline interval) and with an offset not until after 1500 ms. In the 2004 study, in addition to the reported "delayed ELAN" between 300 and 500 ms (relative to word onset) at FT7, there also appears to have been a very early left anterior negativity around 100 ms at electrode F3 (that is, right after the offset

[8] Note that it may be controversial whether the 20% vs. 80% manipulation really leads to distinct (conscious) processing *strategies* or rather involves (subconscious) implicit/statistical learning (see, e.g., Perruchet & Pacton, 2006). For simplicity and consistency of argument, in this paper we will adopt Hahne and Friederici's (1999) interpretation in terms of strategic effects. We thank an anonymous reviewer for helpful discussion on this point.

[9] Note that whereas in German reading studies the word initial letter distinguishes nouns (upper-case) from verbs (lower case), this distinction does not play a role in the two auditory studies discussed here.

[10] A quick count of words starting with either 'ge-' or 'be-' in the web-based CELEX data base (German Lemmas) revealed 943 'ge-' and 1313 'be-' entries, the majority of which were nouns (>600 and >500, respectively) (http://celex.mpi.nl/). Of the 1313 'be-' entries, 544 were verbs (most of which can be assumed to also form past participles starting with 'be-'). Many of the nouns had a high frequency of occurrence. In other words, 'ge-' and 'be-' cannot be viewed as particularly specific to the word category of verbs.

[11] This hypothesis is easy to test: (1) One would expect the effects to be absent during the first trials of a study. (2) Including nouns beginning with the critical prefix should either prevent early effects on verbs violating the phrase structure OR the (grammatical) nouns should also elicit an ELAN. Any of these effects would be highly problematic for the model (again underlining its high degree of falsifiability).

of the 0–100 ms post onset baseline; see Fig. 1 in Friederici et al., 2004; Electrode F3 in *both* the syntactic and the double violation). As none of these early negativities were discussed by the authors, their status must remain speculative (but see Section 2.2.7).

However, there were two other studies (Rossi et al., 2005, Rossi, Gugler, Friederici, & Hahne, 2006), which also used word-final suffixes to mark the WC and did report ELANs relative to the *word onset* (and not relative to the disambiguating suffix), as illustrated in example (4).

(4)	a.	*Der Junge im ****singt*** *ein Lied.*
		(The boy in-the <u>sings</u> a song.)
	b.	*Der Junge im Kindergarten* ***singt*** *ein Lied.*
		The boy in-the kindergarten <u>sings</u> a song.

As in previous studies, the expected noun was replaced by a verb, but this time in sentences with active (rather than passive) voice. On average the verb duration in this auditory study was 629 ms.[12] Given this, it is *not possible* that the critical word-final suffix (−**t**) signaling the WC violation could have been encountered prior to 200 ms. Nevertheless, this study, in sharp contrast to the 1996 and 2004 studies, did report ELAN effects with an onset between 100 and 200 ms relative to *word* onset. [13] Interestingly, in a replication of this study in late second language learners (Rossi et al., 2006), all groups of participants – including both German and Italian participants at low levels of L2 proficiency – displayed the same early onset ELAN/sustained negativity as the native speakers in the 2005 study.[14] None of these data support the claim of a latency shift in ELAN effects tied to the temporal availability of word category information, as the onset of the effects occurred *prior to* the point in time where this information actually manifested in the incoming speech stream. To the best of our knowledge, such early effects can only be explained in terms of either acoustic/phonological (e.g., prosodic) differences on the target word (which could be argued to have been ruled out due to the cross-splicing technique used), *or* by the kinds of context/baseline problems we discuss next.

2.2. Problem #2: methodological issues (beyond the face value)

One might ask regarding potential methodological problems in the ELAN literature: how serious *could* they be? That is, even if some studies used suboptimal designs or analyses, is it not extremely unlikely that *all* of the studies in the ELAN literature are methodologically problematic? Our answer to this is that the vast majority of studies reporting ELANs have actually used only a very limited range of paradigms. Thus, if these designs are systematically flawed, then it is entirely reasonable to question the corresponding body of empirical data.

2.2.1. Two types of PS-violation paradigms

There are two general approaches to generating PS-violation paradigms in ERP research: (A) TARGET MANIPULATION and (B) CONTEXT MANIPULATION, the latter of which we have already briefly introduced (see (1)/(2)). As we will see, ERP data for PS violations appear to covary with the type of experimental approach.

Only a small number of studies have pursued the target manipulation strategy (see Appendix), which keeps the *context constant* while varying the category of the critical words. Thus, in the Dutch example in (5), Hagoort et al., 2003, added the verbal suffix *-t* in (5a).

(5)	a.	*De houthakker ontweek de ijdele ****schroeft*** *op dinsdag.*
		The lumberjack dodged the vain <u>propelled</u> on Tuesday.
	b.	*De houthakker ontweek de ijdele* ***schroef*** *op dinsdag.*
		The lumberjack dodged the vain <u>propeller</u> on Tuesday.

Although this example, just like the German paradigm in (2), also involves replacing an expected noun by a verb, the reading study in (5) found anterior negativities between 300 and 550 ms. This is rather later than typical ELAN findings in other reading studies which – when they have been found at all – have been reported to occur within less than 250 ms post-target-word onset. Interestingly, there is not a single PS violation study that used this type of design and reported negativities in the typical ELAN time range, either in reading or in the auditory modality.

An inherent potential problem in target-manipulation studies is that care must be taken to ensure that genuine *violation* effects can be teased apart from any effects that may be tied to the target word differences themselves (e.g., *lexical* differences between nouns and verbs that may occur independently of the violation).[15]

In contrast to the study above, the majority of PS violation studies have employed the CONTEXT MANIPULATION approach (all unshaded rows in Tables A1 and A2, Appendix), which keeps the target words *constant*, thus avoiding confounds of violation effects and lexical differences. However, in these paradigms, for example in the German paradigm (2), the words immediately preceding the critical targets *systematically* differ between the violation condition and any of the correct control conditions (i.e., *am* versus *wurde/Montag/oft* in (2a–d). This is problematic because such contrasts may give rise to *context effects* on target word ERPs which are independent of the PS violation that such experimental designs are introduced to investigate. For example, if violation condition (2a) is compared to control conditions (2c) or (2d), the content words *Montag* in (2c) and *oft* in (2d) are likely to elicit a larger N400 between 300 and 500 ms compared to the preposition *am* in (2a). In most ERP reading studies, sentences are presented word-by-word, at a rate of one word each 500 ms. Thus, a standard pre-stimulus baseline interval of −100 to 0 ms relative to the onset of the target word *gebügelt* would pick up such N400 word differences and cre-

[12] This information was provided for the (additional) agreement violation condition only, which employed the same verb stems with a slightly more complex inflectional morpheme (e.g., *sing-<u>st</u>*). Given that this suffix (*-st*) is unlikely to account for more than 230 ms, it seems safe to assume a verb stem duration in the PS violation condition of at some 400 ms. Thus word category information became available only around 400 ms, which is still 200 ms later than the onset latency of the reported ELAN.

[13] Ironically, the stimuli of the violation condition were created using a digital cross-splicing technique that highlights the problem at hand: "The category and combined violation were spoken as complete correct sentences including a noun of the prepositional phrase (Der Junge i*m* "Singrau*m*" singt ein Lied. – English: The boy in-the singing-room sings a song) in order to avoid possible phonological influences (...). This additional noun was afterwards excised from the acoustic file" (Rossi et al., 2005, p. 229). Note that the authors' description of this procedure explicitly mentions a *grammatical* sentence that was phonetically identical to the violation condition up to the disambiguating verb-final suffix (i.e., *"Der Junge im Sing[raum]..."*), implying that the ELAN observed at verb onset must either be due to an artifact in sound editing or to some other context effect.

[14] In contrast, for agreement violations low proficiency L2 learners did not show the LAN component found in native speakers and high-proficiency L2 learners. This agreement violation condition used target manipulation, i.e., shared the same context with the correct control condition (*Der Junge im Kindergarten ...*) and was, therefore, unlikely to have any baseline problems that could result in offset artifacts (i.e., sustained negativities).

[15] One way of controlling for pure lexical differences is to include control conditions that contain the critical target words in grammatical contexts (see Steinhauer, Drury, Ullman, & Steinhauer (2010) for a discussion on this issue).

ate an ERP artifact, i.e., a difference between the two conditions that shows up on the target word (*gebügelt*) but is not related to the processing of the target word at all.

Additional types of context effects can occur in *auditory* studies due to phonological, especially prosodic, differences (in terms of pitch contour, signal intensity and duration) between violation and control conditions. For example, (i) stress patterns and duration of spoken pre-target words may differ; (ii) the prosodic contour of the pre-target word context may systematically differ; and (iii) if violations were derived by cross-splicing techniques, splicing artifacts or disruptions of co-articulations or prosodic contours may elicit ERP differences. Moreover, (iv) while computer-controlled visual presentation in reading studies ensures that at least the *target word* stimuli employed in context-manipulation approaches are kept constant across conditions (*by definition*), in spoken sentences even lexically identical target words may still differ prosodically (e.g., in duration of the speech signal or its intonation).[16]

In sum, in experiments relying on a *context manipulation* approach to word category violations, many factors other than the syntactic violation of interest can – in principle – contribute to ERP differences observed on target words (e.g., see also note [4], Table A2, Appendix). While some of these concerns hold equally for visual and auditory studies, there are additional concerns tied to auditory studies in particular. In fact, recent advances in ERP research investigating prosodic effects provide strong evidence that factors such as word stress, word duration and prosodic boundaries are systematically reflected by ERPs, including both negative and positive-going waves (e.g., Eckstein & Friederici, 2006; Pauker, Itzhak, Baum, & Steinhauer, 2011; Steinhauer, Alter, & Friederici, 1999).

2.2.2. A note on (a)symmetry in violation paradigm designs

Before we turn to discuss the issue of context effects relative to published studies, it should be noted that, although lexical and context effects can affect any kind of psycholinguistic ERP comparison, PS violation studies are particularly prone to them as the overwhelming majority of studies have used asymmetric context-manipulating designs. In contrast, lexical-semantic N400 studies and morpho-syntactic studies reporting LANs and P600s usually employ symmetrical (balanced) designs, in which both the context and the target word manipulation are counter-balanced between violation and control conditions. For example, the subject-verb agreement violations in (6) are created in singular and plural forms for half of the sentences, respectively, such that across trials, the exact same words contribute equally to violation as well as control conditions (with respect to both the contexts (The boys, the boy) and the target words (play, plays)).

(6)	a.	*The boys* ***play*** *a game.*
	b.	*The boy* ***plays*** *a game.*
	c.	*The boys ****plays*** *a game.*
	d.	*The boy ****play*** *a game.*

This kind of balanced design avoids virtually all problems that are due to both lexical and contextual differences (except for a few potential prosodic confounds in auditory studies), hence such paradigms typically do not appear to have baseline problems even when tested along with PS violations that do seem to have such problems (e.g., Friederici et al., 1993; Friederici et al., 1999; Hinojosa et al., 2003; Newman et al., 2007; Rossi et al., 2005, 2006).

The fact that balanced designs have not been used in previous word category violation studies does not primarily reflect a lack of care when designing the experiments. In some languages (e.g., in German), even creating a single (i.e., asymmetrical) PS violation that meets the ERP requirement of rendering the sentence unambiguously ungrammatical on the target word is not a trivial task (see Section 2.1.2 on 'outright' violations). To give an example, even a sequence of five German function words in a row, such as *"der mit zum meine von"* (English: *the with for-the my by*) can, in fact, be integrated in an awkward but entirely grammatical sentence.[17] Exactly at which word the brain would encounter processing problems is not easy to predict. Moreover, the difficulties likely have to do with limited working memory capacity rather than linguistic anomalies (as with standard cases of multiply center-embedded structures in English, for example – see Miller & Chomsky, 1963). In many other cases, an apparent word category violation can be reinterpreted as either a morpho-syntactic or conceptual semantic anomaly, neither of which are expected to elicit ELANs.

2.2.3. Context effects in ERPs: spillover and offset

Context effects, as we will understand this notion here, correspond to any kind of ERP effect measured at critical target words which are actually due to upstream processing differences across conditions. Such artifacts typically manifest as early components and are likely to occur in context-manipulating designs, which is important because all ELAN findings at word onset have exclusively been observed in context-manipulating designs. We can distinguish between two different kinds of context effects: spillover effects and DC offset artifacts.

A *spillover effect* occurs when ERP differences associated with words prior to the critical/target words show up after onset of the targets. These could in principle take the form of either: (i) effects which already manifest on the prior word but which have a sustained time-course into the post-target-onset latency ranges or (ii) effects which are tied to the processing of the previous word but simply happen to be late in onset. For example, a sustained negative shift (such as an N400-700; Brown, Hagoort, & ter Keurs, 1999; Neville, Mills, & Lawson, 1992) beginning on pre-target words which gradually increases in amplitude over time (i.e., any "ramp-like" shifts) could in principle survive ERP baseline correction and manifest as a "downstream" early negativity for target words (i.e., right after the baseline interval; see Fig. 2A below for an illustration). For another example, P600s are known to manifest late (after 500 ms), so these effects are typically not seen until after the onset of the next word. ERPs time-locked to this next word with a typical pre-stimulus baseline would show a positivity with early onset because it was triggered by the previous word. The point is that it is possible that a late effect elicited by words immediately preceding critical/target words could be mistaken for an early target word effect.

[16] Prosodic differences on target words can be expected for auditory experiments whose violation conditions were either recorded separately (e.g., Friederici et al., 1993) or that were derived by means of cross-splicing but used speech signals other than those used as actual controls in the relevant ERP contrasts (e.g., Rossi et al., 2005).

[17] This is illustrated by the following example. Note that constituents between matching brackets are adjuncts or optional modifiers whose deletion does not affect the grammaticality of the sentence. [***Der*** [***mit*** [***zum*** [***meine*** [***von*** *frueheren Ereignissen sowieso schon angegriffenen*] *Nerven strapazierenden*] *hundertsten Male*] *aus der Tasche haengendem Taschentuch*] *zur Arbeit erscheinende junge Mann*] *nahm Platz und schwieg.* **Paraphrase: The** young man who came to work, **with** his handkerchief hanging out of his pocket **for** the hundredth time (which really grated on **my** nerves, which had already been ruined **by** previous events anyway) took a seat and remained silent.

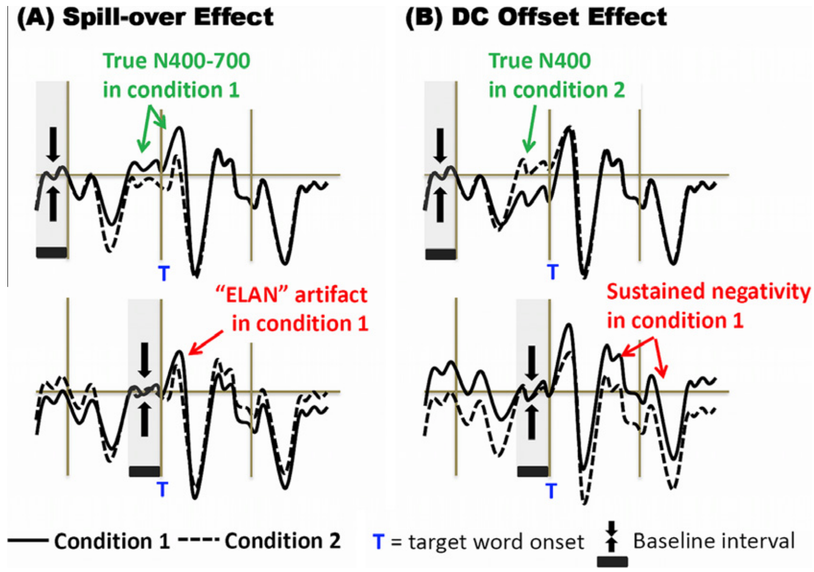

Fig. 2. Schematic illustration of *spill-over* (A) and *offset* (B) effects (not real data). Grey vertical bars mark onsets of words in RSVP, and onset of a given target word is marked by T. Both (A)/(B) show alternative baseline corrections for T, either in the interval immediately preceding T (lower panel), or in the interval immediately preceding the prior word (upper panel). The baseline correction procedure forces the ERPs of all conditions not to differ from each other in the baseline interval. The upper left panel (**A**) shows a hypothetical N400-700 effect (typically left/anterior in distribution), assuming in Condition 1 the target (T) is preceded by a function word, and in Condition 2 that same target word is preceded by a content word. This effect "spills-over" into the target word latency ranges, and the nature of this effect is clear if we inspect ERPs for the word prior to T (upper panel of A). However, ignoring this difference, and only inspecting the target word latency ranges (as in the lower panel of (A)), we see that the N400-700 effect could survive baseline correction, yielding the appearance of an early ELAN-like effect in Condition 1, time-locked to T. In the right panels (**B**), we see a hypothetical N400 for Condition 2 in the baseline interval (upper panel), but given the alternative baseline correction (lower panel) the entire waveform for Condition 2 would be positively shifted. As a consequence, the type of artifact in this case would be a sustained positive shift in Condition 2, which however is indistinguishable from a *sustained negative shift* in the violation condition 1. Such sustained shifts are referred to as DC offset effects (artifacts), or simply as offset effects.

In contrast to spillover effects, *offset effects* would be cases where differences in the baseline interval introduce spurious shifts in downstream ERPs. Thus, if the experimental condition is more *positive* than the control condition during the baseline interval, the ERP baseline correction compensates for this difference and shifts the entire ERP curve into the negative amplitude range,[18] resulting in a *sustained negativity* (see Fig. 2B for an illustration).

These two types of possible context artifacts in ERPs are of relevance because ELAN-like effects in context-manipulating PS violation studies such as those in (1) and (2) have actually also manifested in two distinct ways, with some studies demonstrating a *local/transient effect* (usually peaking around 150 ms), while others seemed to elicit early negativities that "merged into an *anterior sustained wave*" (Ye, Luo, Friederici, & Zhou, 2006, p. 190; see also Isel, Hahne, Maess, & Friederici, 2007; Rossi et al., 2005, 2006), usually lasting for several hundreds of milliseconds. We will argue that these two data patterns may be at least partly connected to the two types of context effects (artifacts).

Though some types of context effects can arise in essentially the same ways across visual and auditory ERP studies, some considerations in connection with the latter deserve special attention. Accordingly, we will first discuss context/baseline issues in reading and then turn to a discussion of these issues in the auditory domain.

2.2.4. Context/baseline problems in ELAN reading studies

Without exception, published reading studies which have shown ELAN-like effects before 250 ms post-stimulus onset have involved contrasts of unbalanced stimuli (see Appendix, Table A1).[19] The nature of these unbalanced context manipulation paradigms suggests alternative accounts of ELAN effects in terms of pre-target word differences (e.g., word class). That word class differences preceding the targets could influence downstream ERPs via spillover or offset effects is not in doubt. What is not known is whether confounds of this sort may generally account for ELAN findings in reading studies.

One study which systematically examined ERP effects for words preceding the critical target words – in fact, the only study to have done so – was Friederici et al. (1999). There, a marginal ELAN effect was found at a single electrode (F7), which was significant only in a narrow 100–150 ms time-window (and only using a −50 to +50 ms pre-to-post-critical word baseline correction). However, robust effects tied to elements upstream of the critical/target words were also documented in that study, with the words immediately preceding the targets (an adverb versus a preposition, as in (2a/d) above) exhibiting a word-class related N400 effect (with the adverb more negative-going). This was explicitly argued to have resulted in a reduction of the ELAN effect on the target word comparison (see their pages 444, 447), and indeed this is quite plausible. A larger N400 for the words preceding the targets in (2d) relative to (2a) could have partially canceled out an early negativity in the other direction for the subsequent critical target verbs. That the ELAN thus might only survive as a significant effect

[18] Note that, all else equal, such shifts could be in principle expected to have arbitrarily long durations. However, in practice the duration of such shifts would, crucially, also depend on filter characteristics used in data pre-processing. That is, any high-pass or band-pass filter will ultimately limit their duration. Further, the duration of such effects may also be influenced by other positive/negative-going shifts arising in virtue of downstream processes.

[19] An anonymous reviewer points to Roehm & Haider (2009) as an example of a balanced study design showing ELAN-like effects for PS violations without baseline problems. However, the study did not use a balanced design of the kind shown in (6). Moreover, although a frontal effect between 270 and 370 ms in their Experiment 1 formally met the criteria for an ELAN, the *posterior* negativity between 340 and 440 ms found for the same type of violation in Experiment 2 resembled an N400. Thus, although this study did not seem to show any context-driven artifacts, we are not convinced the effects can be viewed as ELANs. Certainly, except for the first 30 ms of the analysis interval, even the timing of the Experiment 1 effect is consistent with processes occurring *after* Friederici's Phase-1.

at one left anterior scalp electrode is clearly consistent with this scenario.

However, the existence of such spillover effects from prior words also runs the other way, perhaps accounting for the fact that an ELAN was found at all. Given that function words also are known to elicit relative negativities compared to content words over left anterior scalp regions (N400-700), the presence of the marginal ELAN may itself constitute such a spill-over effect. Inspection of Friederici et al.'s (1999) data suggest this may have been the case: at electrode F7 there is a relative negativity for their syntactic violations prior to the onset of the critical/target word (see their Fig. 1). Moreover, this apparent pre-target-word effect does not manifest clearly at any other electrode (and neither does the ELAN).

Context issues associated with a version of the Neville et al. (1991) paradigm are explicitly discussed in Newman et al. (2007), where contrasts in (6a/b) were tested:

(6)	a.	*Yesterday I drank his* brandy ***by*** *the fire.*
	b.	*Yesterday I drank* his *****by*** *brandy the fire.*

Newman et al. report that after data from half of their participants had been collected, a preliminary inspection of the ERPs associated with this contrast showed differences already at onset of the target word by (i.e., at 0 ms). They assumed that this was due to ERP differences associated with the word class distinctions realized on the prior word (e.g., *brandy* versus *his* in (6a/b), respectively). At that point in their data collection, they changed the stimuli to more closely resemble (1a/b) by switching out the pronouns for proper names. However, as can be clearly seen in their grand average data, baseline issues nonetheless remained (i.e., ERP differences still emerge at target word onset – i.e., at 0 ms – see their Fig. 1b). As the authors acknowledge, this makes it impossible to determine whether their experiment elicited an ELAN effect independent of differences attributable to prior words.

Though baseline issues are quite typically not carefully addressed in the ELAN reading literature, even where they are ignored it is sometimes possible to see clear indications of such problems. For example, Hinojosa et al. (2003) set out to compare (E/)LAN effects for morpho-syntactic and phrase-structure violations directly, using the Spanish paradigm in (7) (i.e., derived from the German paradigm in (2)). Here the pre-target items in the phrase-structure violation condition (7b) were function/closed-class words (i.e., determiners) while both the correct control (7a) and the morpho-syntactic violation (7c) contained content/open-class words in the pre-target position.

(7)	a.	*La prueba ocultada por el* fiscal ***apareci***ó.
		The proof (that was) hidden by the public prosecutor (it) appeared.
	b.	*La prueba ocultada por* el ****apareci***ó.
		The proof (that was) hidden by the appeared.
	c.	*La prueba ocultada por el* fiscal * ***apareci***í.
		The proof (that was) hidden by the public prosecutor (I) appeared.

The potential importance of this study derives from the possibility of directly comparing (E)LAN-type effects for the (7b) and (7c) violation types. They reported three main findings: (i) both violation types elicited comparable "LAN" effects (250–400 ms), (ii) both violations elicited subsequent P600 effects which were attenuated for the PS relative to the morpho-syntactic violations, and (iii) the PS violation also elicited a posterior N400 effect.

However, in this study it is clear based on visual inspection of their ERP data that pre-target word differences played a major role in their reported findings. Hinojosa et al. used a −200 to 0 ms baseline interval which can be seen in their data to have picked up (pre-target) N400 differences. As a consequence, the baseline correction clearly shifted the entire waveform of the phrase-structure violation condition, relative to the other two conditions, towards the negative amplitude range (see our Fig. 2B). Note that this kind of problem is typically reflected by the appearance of a cross-over of the contrasted ERP waves during (or prior to) the baseline interval (e.g., a positive slope in one condition and a negative slope in the other; this was the case in this particular study, see Hinojosa et al.'s Fig. 2). Alternative baseline corrections (e.g., either −100 to 0 ms or a post-target 0–100 ms baseline) would easily yield a qualitatively different ERP profile for the target words. Specifically, it appears that none of the negative-going effects for the phrase-structure violation condition would have survived such alternative (and more appropriate) baseline corrections, and that instead of finding a smaller P600 effect for the PS-violation there would rather have been a larger such effect. Thus we believe the data cannot be interpreted in the way presented by the authors.

This represents an instance of what we referred to above as an *offset effect*, that is: starting no later than the onset of the target word (i.e., at 0 ms) the entire waveform was shifted into the more negative range (yielding their observed pattern). Finally, note that the comparisons which employed *target* manipulation and kept the context constant (i.e., the morpho-syntactic violation contrast in (7c) *versus* (7a)) also did not show any indication of such baseline problems.

But what about the initial Neville et al. (1991) paradigm (1a/b), or the following contrast used in Yamada & Neville (2007)?

(8)	a.	*Mommy can cut the meat* with ***that*** *knife.*
	b.	*Mommy can cut the meat with* her ****that*** *knife.*

It is less obvious that pre-target word differences could have played a role as these were function words in both (8a) and (8b), and proper names vs. common nouns in Neville et al. (1991). However, the only way to be certain that contextual differences upstream of the critical target word did not infect the target word ERPs is to actually verify this directly by examining pre-target word ERPs (as in Friederici et al., 1999), and no such demonstration has been offered.

Data from two recent studies (Drury et al., 2010; Steinhauer, Drury, Portner, Walenski, & Ullman, 2010) demonstrate that this general concern about context manipulation is not merely speculative. These studies included violations modeled closely after the original Neville et al. (1991) paradigm (i.e., see (1a/b) above). Our findings for this same paradigm were very similar to those of the original Neville et al. study. But, our data were also subjected to additional analyses on ERPs time-locked "one word back" from the target prepositions (Drury et al., 2010), which demonstrated robust differences for the immediately preceding pre-target words. ERPs for the proper names (e.g., *Max's*) in the violation condition elicited a sustained left anterior negative-going shift relative to the common nouns (e.g., *proof* in the control condition). And crucially these effects tended to increase in amplitude, spilled-over into the target word ERPs, and overlapped with the "downstream" ELAN effect entirely.

Note, of course, that it remains an unaddressed open empirical question whether baseline problems of the sort that we have found to be connected with the original Neville et al. (1991) paradigm also emerge in other similar context manipulation paradigms (e.g., the (8)a/b contrast used in Yamada & Neville (2007)). Our suspicion is that these problems are quite general. At the very least, however, we argue that in order to demonstrate a "real" ELAN effect using *any such unbalanced context manipulation paradigms*, the possibility of confounds in the baseline interval must be rigorously ruled out. Furthermore, we submit this is an analytical burden that no single existing prior reading study reporting ELAN-like effects has succeeded in discharging (see also note [5], Table A1, Appendix).

2.2.5. Context/baseline problems in auditory ELAN studies

Where ELAN-like effects have been shown in reading studies, these have manifested as local (i.e., transient), sometimes extremely short-lived effects within the first quarter second post-target-word onset. Given that the first phase of Friederici's model is assumed to last from 100 to 300 ms, this pattern (if real) is exactly what one would expect for a component alleged to index the action of a fast, automatic encapsulated processing sub-system.

In the auditory domain however, effects referred to as ELAN components actually come in two different flavors which appear to be more-or-less equally frequent in the literature. The first is a local effect between 100 and 300 ms and is thus quite similar to its counterpart in the visual domain. The second, however, although similar in scalp topography (i.e., at frontal electrodes) displays a similarly early onset but lasts up to more than 1000 ms, thus enjoying a duration which spans not only the first but both of the subsequent two phases of the model as well.

Given that processes occurring even after only 400 ms are alleged to be *controlled*, and not *automatic*, a crucial question is whether these sustained negativities actually reflect just one long-lasting process or rather two (or more) distinct processes which just happen to share a very similar scalp distribution. According to Friederici and colleagues, these sustained effects are to be understood as (i) ELANs (the early part of such sustained shifts) plus (ii) something else (i.e., the rest of these shifts, perhaps working memory related; Friederici, 2002). This interpretation appears to be partly motivated by the frequent presence of an early local negative peak around 150 ms (identified as the peak latency of the ELAN). One problem with this interpretation is that often the corresponding control condition *also* demonstrates a peak in the same time range (see Fig. 3 below) such that the *difference wave* reflecting the actual effect does not necessarily show such a local peak. The functional difference between just local effects and such sustained negativities is still unknown.[20] For the moment, we will set this issue to the side and simply assume the prevailing view that there is in fact a unitary "ELAN" effect which underlies what – at least superficially – looks like two distinct types of ERP responses.

As mentioned above, two studies which reported sustained negativities were those of Rossi et al. (2005, 2006) where, remarkably, such effects manifested in two experimental conditions across five different populations (including low proficiency second language learners). This finding is even more remarkable given that the onset of these sustained negativities preceded the crucial word category information. Since these effects clearly cannot be diagnosed as brain responses to syntactic violations, what *are* they?

A satisfactory answer to this question may potentially provide a general account for similar findings in other studies. Recall that all studies employing the prefixes *ge-/be-* (Friederici et al., 1993; Hahne & Friederici, 1999, 2002; Sabisch, Hahne, Glass, von Suchodoletz, & Friederici, 2006) have in principle the same problem as the Rossi et al. studies in that these prefixes do not provide reliable word category information to cause the kind of "outright" violation assumed to drive the reported ELAN effects. Crucially, in order to be able to evaluate the validity of such concerns one needs to apply the same analytical strategy discussed above, namely careful inspection of pre-target word latency ranges.

Unfortunately, however, the vast majority of auditory studies that are particularly prone to these concerns (e.g., those employing the German context-manipulation approach) have used somewhat unusual 0–100 ms baselines and generally display the ERPs only *after* onset of the target word. Thus, for such studies it is not possible to determine (as could be done, e.g., for the Hinojosa et al., 2003 reading study) whether either spillover or offset effects may have contributed to the reported target word (violation) effects.

A likely possibility, in our view, is that *offset effects* could be responsible for the long-lasting shifts. Such offset effects *necessarily* result in long-lasting ERP differences between conditions, because the entire brain wave is shifted (see Fig. 2B). The Rossi et al. (2005, 2006) studies are the most obvious examples as their early negativities cannot be accounted for by word category violations. In these studies, like in the Hinojosa et al. (2003) reading study, the pre-target element was a function word in the PS violation condition (e.g., *im* in example (4a) above), whereas it was a content word the control condition (e.g., *Kindergarten* in example (4b)). Thus, during their baseline interval one can assume an N400 to be present in the control condition, such that the baseline correction is likely to have shifted the control condition toward the positive amplitude range, resulting in a sustained negativity in the violation condition. One might argue that N400s do not typically have left-anterior distributions, so how could this effect have resulted in an ELAN-like effect? The answer is that it did not. Closer inspection of the actual distribution of the sustained negativity in Rossi et al. (2005) reveals that the negativity had a broad *bilateral* distribution and was most significant at *central* sites (both at midline and lateral electrodes), in line with N400 distributions in many studies. Its interpretation as an *anterior* negativity by Rossi and colleagues (especially in the 2006 paper) was justified by the fact that it was more significant at anterior *compared to posterior electrodes*, whereas the lack of left-lateralization was found to be in line with other 'ELAN' findings that also turned out not to be left-lateralized. It is also worth mentioning that the authors' distinction between an early (100–200 ms) ELAN part and a later negativity (200–600 ms) is not justified by the data at hand, which simply show one single (sustained) negativity with early onset, the broad distribution of which was identical in both analysis windows. Intriguingly, this sustained negativity was present at F7 and F8 from 100 ms (i.e., right after the baseline interval) and lasted until the end of the time epoch shown (i.e., 1500 ms), which is exactly the pattern expected for DC offset artifacts. At more posterior electrodes and at electrodes closer to the midline, the negativity disappeared between 600 and 1200 ms, but this was obviously due to the large centro-parietal P600 superimposing the negativity in this interval.[21] Whereas the sustained negativity in Rossi et al. (2005, 2006) seems to be at least partly influenced by word class N400 effects in the baseline interval, we do not believe all offset artifacts are due to word class differences. For example, other lexical or *prosodic* differences in the baseline interval may result in offset components with distinct topographies, including those with a more frontal prominence (Hwang & Steinhauer, 2010, in press, in preparation; Itz-

[20] Pakulak & Neville (2010) suggested that early (proficiency-dependent) effects between 100 and 300 ms may then affect subsequent processes reflected by late negativities and P600s. Below, we will present an alternative view.

[21] Similar baseline problems may also have played a role in the Chinese PS violation by Ye et al. (2006) whose frontal sustained negativities showed an onset around 50 ms when using a rather unusual −1100 to 0 ms baseline. They only report a negativity in absence of a P600, but a baseline offset may actually have moved the violation waveforms so far into the negative amplitude range that the parietal P600-like positive shift visible in their data simply may not have been large enough to reach significance in addition to cancelling out the negativity.

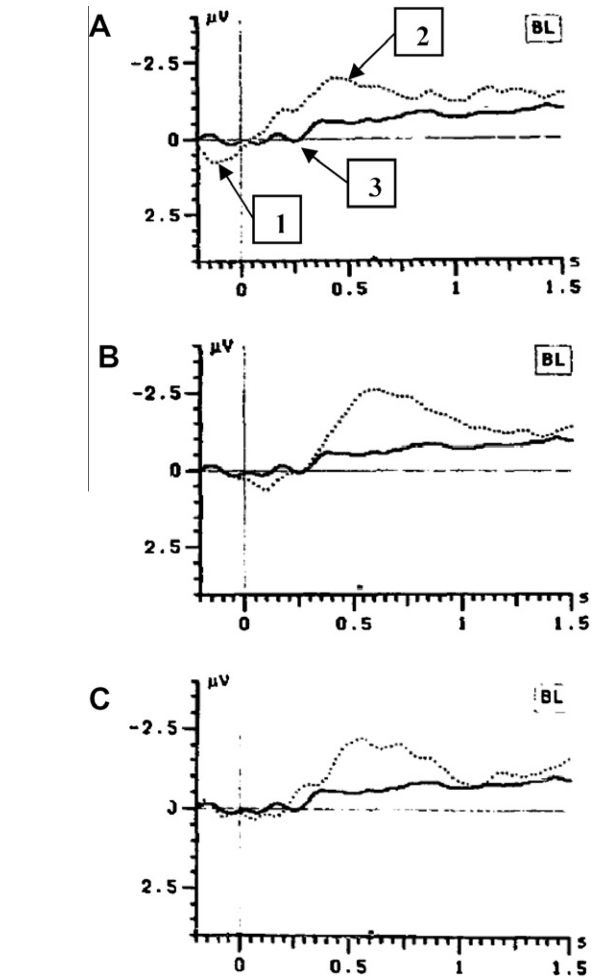

Fig. 3. ERP plots from Friederici et al. (1993, adopted from their Figs. 2-4) at left-anterior electrode "BL", time-locked to target word onset. The three plots illustrate (**A**) word category violations, (**B**) semantic anomalies and (**C**) morpho-syntactic violations (all represented by dotted lines) as compared to the correct control (solid lines). Note that only the word category violation in (A) displays a difference between conditions prior to target word onset (arrow 1). Arrow 2 indicates the negative maximum of the sustained negativity, while arrow 3 points to a local positive maximum around 250 ms which was sometimes used to motivate the split into an early ELAN time window and a late time window for the subsequent sustained negativity (although *not* in the 1993 study). Note also that this local maximum is present in the control condition as well. **Fig. 3**: Reprinted from, Friederici, Pfeifer, Hahne, with permission from Elsevier.

hak, Pauker, Drury, Baum, & Steinhauer, 2010). Our point here is that offset artifacts may provide a more general account of the prevalence of such sustained negativities that start right after the baseline interval (typically 0–100 ms or 0–200 ms *post* target word onset).

In fact, the specific choice of baseline intervals in many such studies may directly contribute to the particular profile of these sustained effects. Given that – by definition – ERP effects are prevented from showing up *during* the baseline interval, the selection of a 0–100 ms baseline interval would actually *ensure* that the negativities start in the 100–200 ms time window, i.e., exactly when ELANs are supposed to occur. The only way one could tell whether this was in fact the case would be to look at the time interval prior to the baseline interval. Especially if the experimental condition is initially more *positive* than its control, an offset contribution to the negativities is very likely (as illustrated in Fig. 2B). Unfortunately, as noted, most auditory ELAN studies simply do not provide this information as ERP plots are presented starting only at target onset (Friederici et al., 1996; Hahne & Friederici, 1999, 2002; Hahne & Jescheniak, 2001; Rossi et al., 2005, 2006; Sabisch et al., 2006; Mueller et al., 2005).

However, there is one notable exception: the very first study reporting ELAN effects elicited by the German paradigm (Friederici et al., 1993). There, unlike in subsequent work, a *distant* baseline interval of 250 ms relative to the auxiliary "*wurde*" (see example (2a/b) above) was selected, and plots time-locked to the target past participle are shown *including* a 200 ms pre-stimulus interval (i.e., −200 to 0 ms). What this plot illustrates (see our Fig. 3) is that at the single left-anterior electrode depicted ("BL" for Broca's area in the Left hemisphere), the violation condition is initially more *positive* than the control and has a local positive maximum (∼1 μV) at about −150 ms (arrow 1). It then starts to display a slow negative shift that crosses both the baseline and the ERP curve of the control condition and reaches its negative peak (∼2 μV) around 500 ms (arrow 2). Arrow 3 indicates a small local positive peak which is preceded by a small local negative peak in the control condition. The corresponding pattern in the violation condition is sometimes used to identify the ELAN peak and to separate the early 'ELAN' portion from the subsequent 'late negativity'.

What is the significance of this apparent pattern? The point is that a baseline correction using *any* interval between the positive peak (−150 ms) and the subsequent negative peak (+500 ms) would produce a sustained negativity in the target word epoch, the onset of which would necessarily be aligned to the end of that baseline interval. Setting to the side the obvious question of what, precisely, may account for this pre-target positivity, the important point is that the sentence contexts used in this case are the same as in most other auditory ELAN studies using the German paradigm. And, given that this positivity was most prominent at this left anterior electrode, the 0–100 ms (post-target) baselines used in all auditory ELAN studies (after 1996) employing the German paradigm would – if this effect is general – have *created* sustained left anterior negativities with onsets at 100 ms. It is also important to note that this argument is *not* undermined by studies reporting that a standard *pre-stimulus* baseline interval (−100 to 0 ms) did not change the relevant effects (e.g., Hahne & Friederici, 1999, 2002).[22]

2.2.6. Local ELANs, sustained negativities, and P600s

If the foregoing is correct, how does this relate to the findings showing local/transient ELAN effects, instead of the sustained negative shifts? Recall that ELANs are usually followed by large P600 components. Thus, one logical possibility is that the local effects are actually the result of the superposition of sustained negativities and concurrent P600-type effects. In fact, a careful inspection of the available ERP data showing local effects suggests that their offsets very often coincide with the onset of such posterior positivi-

[22] The reason is as follows: A *pre-stimulus* baseline interval (e.g., −100 to 0 ms) would only marginally change this data pattern. It would primarily shift the ERP curve of the violation condition even further into the negative amplitude range, thereby (i) increasing its overall (negative) amplitude and (ii) reducing the onset latency of the 'violation' negativity. The first point will ensure that in the time windows used for analysis (e.g., 100–300 ms), the negativity is still (or even more) significant, while the P600 would not be strongly affected due to its posterior distribution. (In fact, due to the increased frontal negativity, one might even expect a stronger violation x anterior/posterior interaction while posterior P600 effects should remain largely unaffected. As illustrated in Table 2 of the Hahne & Friederici, 2002 paper, this is exactly the pattern that was reported.) The second point is a more critical one, as ERP onset latencies of less than 100 ms are generally suspicious and easy to identify as a context-driven artifact (rather than a real violation effect). However, in order to actually be able to estimate whether such a latency shift of the 'ELAN' would result from a different baseline interval, one would need to (a) *show* the ERP plots prior to target word onset and/or (b) run corresponding statistical analyses in an additional 0–100 ms time window. Both studies by Hahne & Friederici (1999, 2002) only reran the 100–250 and 300–1000 ms time window analyses but did not provide either pre-target ERP plots or data on the 0–100 ms interval.

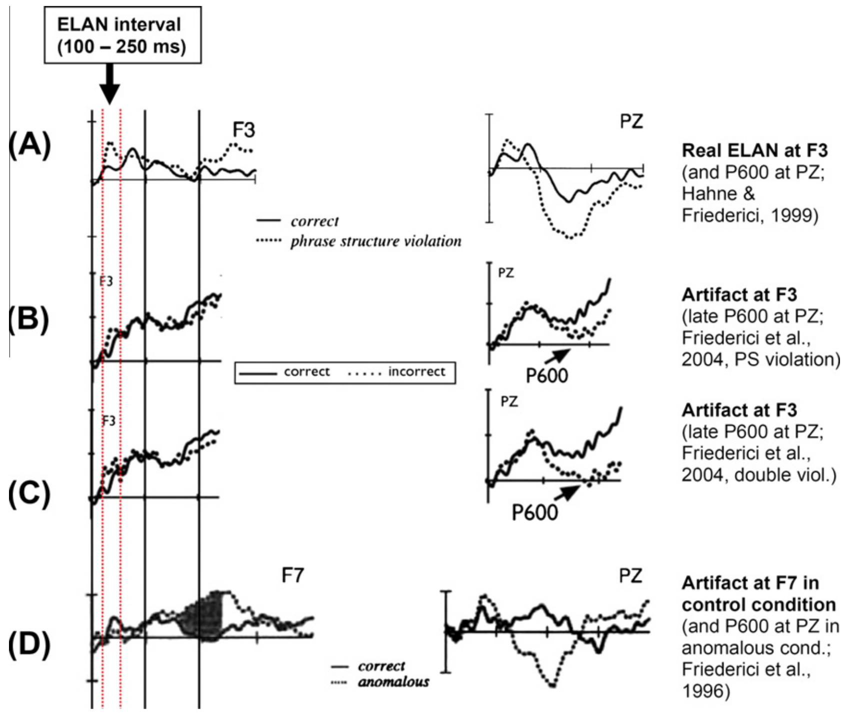

Fig. 4. Direct comparison of (**A**) a 'real' ELAN (from Hahne and Friederici, 1999), ELAN-like artifacts from Friederici et al. (2004) in (**B**) their pure PS and (**C**) in their double violation conditions, as well as (**D**) an ELAN-like artifact for grammatical nouns (from Friederici et al., 1996). Plots in the left panel illustrate these negativities at frontal electrodes and are aligned (i.e., have the same time scale), as illustrated by the solid vertical lines at 0, 500 and 1000 ms. Plots in the right panel illustrate the corresponding ERPs at PZ. The dotted vertical lines in the left panel mark the typical ELAN time interval (100–250 ms). In all four cases, a left-anterior negativity is visible in this time interval for target words that followed a preposition, whereas the respective other condition (preceded by the auxiliary *'wurde'* does not show such a negativity. Only the 'real' ELAN in (**A**) is directly followed by a large posterior P600 in the *same* condition, likely superimposing the negativity. After offset of the P600, an anterior negativity is visible again at F3 (see main text). **Fig. 4A:** Reprinted from, Hahne and Friederici (1999), with permission from MIT Press. **Fig. 4B** and **C:** Reprinted from, Friederici, Gunter, Hahne and Mauth (2004), with permission from Wolters Kluwer Health. **Fig. 4D:** Reprinted from, Friederici, Hahne and Mecklinger (1996).

ties. Conversely, studies lacking P600s tend to display sustained negativities.

In this connection, arguably the most telling data come from auditory studies showing both the "local" and the "sustained" types of early negativities *within the same participants*. This is the case in Hahne & Friederici (1999), where the increased 80% proportion of PS violations in the materials prevented the "controlled" P600-type effects from appearing. As a consequence, the first experiment presenting a low 20% proportion of violations shows a "local" ELAN which begins right after the baseline interval (100 ms) and ends at about 300 ms, i.e., exactly when a large P600 develops at central and parietal electrodes (see our Fig. 4A). In contrast, their second experiment (with the 80% violation proportion) shows a sustained negativity and no P600. Similarly, Hahne & Friederici (2002) demonstrated that a semantic task elicited an N400 instead of a P600 in combined PS and semantic double violations, but not in the pure PS violations (which did show a P600). In absence of any P600, the ELAN in the double violation was (part of) a sustained left anterior negativity whereas the local ELAN in the pure PS violation condition was present only until approximately 500 ms, i.e., when the P600 started to develop.

More recently, an auditory study by Pakulak & Neville (2010) investigated the impact of first language (L1) proficiency and socio-economic status (SES) on ERPs elicited by word category violations. They used the same context-manipulation paradigm as Yamada and Neville's (2007) reading study. The high proficiency group showed a local left-lateralized anterior negativity (100–300 ms), followed by a large and relatively long-lasting P600 between 300 and 1200 ms, which was most prominent at parietal electrodes but also "extended to anterior medial sites" (p. 2735). In contrast, the low proficiency group, whose posterior P600 was significantly smaller, displayed a bilateral sustained frontal negativity between 100 and 1200 ms.[23] In this latter group, the negativity was attenuated only between 300 and 700 ms (i.e., when the P600 was present), at medial electrodes. As a whole, data in both groups can be re-interpreted as sustained negativities that were cancelled out by P600s (temporarily in the low proficiency group and from 300 to 1200 ms in the high proficiency group). What the data seem to illustrate is that posterior P600s systematically propagate to frontal electrodes, such that smaller P600 components leave the sustained negativity largely intact (at least at lateral frontal electrodes), whereas large long-lasting P600s can almost completely cancel out the frontal negativity and result in apparent 'local' ELAN-like effects.

Alternatively, as suggested by Pakulak & Neville (2010), one could argue that proficiency-related distributional differences in the frontal negativity also affected the recruitment of processes reflected in the later P600. However, as smaller P600s have been consistently observed in low (versus high) proficiency L2 learners even in absence of frontal negativities (e.g., Rossi et al., 2006; in all conditions for both German and Italian), we suggest that the most parsimonious account for group differences in the (visible) presence versus absence of sustained negativities at frontal electrodes is that of proficiency-related P600 differences. On the other hand, we do agree with Pakulak & Neville (2010) that the stronger *left-lateralization* of the negativity in highly proficient L1 speakers may indeed be influenced by proficiency, independent of the P600. A more detailed discussion is beyond the scope of this paper (but see Steinhauer, White, & Drury, 2009).

In yet another study, Hahne and Jescheniak (2001) demonstrated early (100–250 ms) local ELAN effects in both regular

[23] At medial electrodes the negativity appeared to emerge already at 0 ms (in the low proficiency group) – but analyses for the 0–100 ms latency range were not reported. Note that our interpretation of the sustained negativity patterns in the Pakulak and Neville's study differs from the authors' own interpretation.

German and Jabberwocky sentences for past participles (and pseudo-words) starting with prefix '*ge-*'. What is nicely shown in their data (especially for the regular sentences) is that the P600-complex actually consisted of an early fronto-central positivity peaking at FCz around 250 ms (likely a P3a) and a subsequent posterior P3b/P600 (on the P3a/b distinction, see Donchin, 1981). It appears that it is primarily the early P3a that is responsible for cancelling out the sustained negativity at frontal sites. Both prior and after this P3a, the sustained negativity shows up in the ERPs, especially over right-anterior sites (first as bilateral ELAN-like effect, then again as a late negativity). Moreover, in a Chinese study by Ye et al. (2006), no significant P600 was found at all, so that – as one would expect – a sustained negativity was reported.

A particularly striking case that reveals additional details about the relationship between local and sustained negativities is the auditory study by Hasting & Kotz (2008). Here, the authors tested both (i) morpho-syntactic subject–verb agreement violations (expected to elicit late LANs in phase-2 according to Friederici's model) and (ii) phrase structure violations (expected to elicit ELANs in phase-1). Instead of sentences, the authors used two-word utterances in which a suffix at the end of the respective second word provided the critical information indicating both the presence/absence as well as the type of violation. Moreover, they conducted two experiments that only differed in the task employed: in Experiment 1 subjects performed a grammaticality judgment task (attended condition), in Experiment 2 they watched a silent movie and were asked to ignore the sentences (unattended condition). Unlike in any previous study, the ERP analysis was time-locked to the onset of the disambiguating suffixes in *both* PS and morpho-syntactic conditions. There were a number of remarkable findings. First, in the attended task condition, both types of violation elicited early negativities between 100 and 300 ms, immediately followed by a P600.[24] Thus, the model's categorical distinction between phase-1 ELANs for PS violations and phase-2 LANs for morpho-syntactic violations was not confirmed (see also Capek et al. (2009) for similar effects in sign language). Moreover, the scalp distribution of the early negativity was broad and not confined to left anterior sites. Most interestingly in the present context, where participants were directed to *not attend* to the auditory stimuli (in Experiment 2), the P600 was successfully suppressed. The resulting ERP pattern in both conditions was a sustained, broadly distributed negativity starting at 100 ms and lasting until some 800 ms after suffix onset.[25] Our point here is that the only ERP difference between Experiments 1 and 2 lies in the presence of a P600 (in Experiment 1), which cancelled out the later part of the sustained negativity. In fact, especially in the agreement violation, all nine electrodes displayed show a biphasic ERP pattern consisting of an early negativity that is immediately followed by a P600, such that the offset of the negativity must be viewed as a function of the onset of the P600.

Given the striking similarity of this pattern across auditory ELAN studies, we believe this to be a general explanation for apparent inconsistencies in the duration of early negativities found in syntactic violation conditions. That is, we hypothesize that PS violations in auditory studies *always* elicit sustained negativities and no local ELANs. Whenever an ELAN looks like a local (transient) effect, this is likely to be due to a concurrent P600 component cancelling out the later part of the negativity. Note that this suggestion is quite similar to an argument put forward by Osterhout and colleagues (Osterhout et al., 2004), who argued that the superposition of N400s and P600s may result in apparent LAN-like components. If our hypothesis is true, there may be no need to account for any local ELAN effects between 100 and 300 ms (as suggested by Friederici's model), but there is need to explain sustained negativities with a remarkably early onset. (In Section 2.2.7 we discuss one potential exception.)

An important question is whether all of these sustained negativities reported in the literature are likely to reflect the same cognitive process. Our current view is that they may not. In the Hasting & Kotz (2008) study, these negativities were observed for both PS and agreement violations. As the authors point out, the finding of an invariably very early negativity starting at 100 ms even in an unattended paradigm (Experiment 2) suggests a rather automatic process which nevertheless seems to last for several hundred milliseconds. Importantly, it was not specific to phrase structure violations but was found for agreement violations as well. It is worth mentioning that the authors used a very sophisticated target-manipulation approach and an appropriate −100 to 0 ms pre-stimulus baseline, thus avoiding any context effects or baseline problems. Therefore, their sustained negativity must in fact have been triggered by the different suffixes. Since they also counterbalanced the use of suffixes across violation and control conditions, any trivial explanation of the negativities in terms of systematic acoustic differences can be ruled out as well. In other words, these data *do* provide strong evidence for very early syntax effects (in line with Friederici's model), which however – once time-locked to the relevant information – are not specific to PS violations (contrary to Friederici's model).

Given these findings from the Hasting & Kotz experiments, the idea suggested above that sustained negativities may *in general* be attributable to offset artifacts cannot be maintained (because the Hasting & Kotz study avoided any of the baseline problems we have been discussing). On the other hand, the issues we have raised about the Rossi et al. (2005, 2006) studies revealed that those sustained negativities *are* best explained in terms of context/baseline effects. Together, these findings leave us with at least two types of short-latency sustained negativities that look quite alike but cannot be explained by a single underlying mechanism: one due to syntactic (or morpho-syntactic) anomalies (e.g., Hasting & Kotz, 2008; see also van den Brink & Hagoort, 2004) and another one due to offset artifacts. The question then is whether the majority of ELAN-like findings in auditory studies belong to the first or the second category.

For most auditory PS violation studies using the German paradigm the data remain completely ambiguous as to whether the sustained negativities (or apparent local effects) are true reflections of syntactic anomalies or due to baseline problems (or both). We have shown above that the prefixes *ge-*/*be-*do not provide unambiguous word category information to signal outright PS violations (e.g., in Friederici et al., 1993; Hahne & Friederici, 1999, 2002; Hahne & Jescheniak, 2001), unless we assume a strong strategic influence on local expectations that develop over the course of an experiment and are, therefore, ecologically invalid. On the other hand, the left-anterior distribution of negativities in some of the studies does not seem to point to word class (i.e., N400s) being exclusively responsible for potential offset effects.

What other kind of differences could lead to artifacts in these studies? We hypothesize that many ELAN-like effects in auditory

[24] Only in the agreement violation condition did the P600 result in an independent (significant) P600 effect between 300 and 800 ms. In the PS violation condition, the P600 was strong enough to cancel out the sustained negativity, but the P600 did not show up as an additional positive-going component. This pattern is similar to that in Ye et al. (2006) and is likely due to (a) similar central distributions of both the sustained negativity and the P600 and (b) the fact that the negativity was still present after 300 ms (and thus 'worked against' a significant P600 effect in the 300–800 ms analysis window selected by the authors).

[25] Note that the authors analyzed two time windows: (i) 100–300 ms and (ii) 300–800 ms, and found a significant interaction between violation and the anterior–posterior axis (pointing to a somewhat frontal distribution) only in the second time window. However, the voltage maps for both early and late time windows (in both violation conditions) display very similar fronto-central effects, such that any distinction between the early and the late part of the sustained negativity appears unjustified in absence of statistical analyses directly contrasting the two time windows. In our view, the most parsimonious account for the observed ERP pattern would assume only one single sustained negativity, which either co-occurred with a P600 (Experiment 1) or not (Experiment 2).

studies may be due to *prosodic* context effects. Here, the prosodic analyses carried out by Sabisch and colleagues (2006) for the PS violation stimuli of the auditory German paradigm seem to provide some answers. Sabisch et al. report that, whereas the cross-splicing technique used to create the German violation conditions successfully avoided interruptions of the co-articulation between spliced words, (i) the splicing point at target word onset still disrupted the prosodic contour, (ii) the target word's pitch maximum differed from that of a grammatical noun in the same position, and (iii) the abrupt fall in fundamental frequency signaled sentence final prosody (which would result in an unacceptable structure even in case of a grammatical noun). In addition, not only did the word categories of elements preceding the target words differ in systematic ways (see example 2a–d above), but their prosodic contours did as well. Among these issues, the lexical and prosodic context differences and the disruption of the prosodic contour at target word onset are most likely to have contributed to the early negativities. A similar conclusion was drawn by Mueller et al. (2005) who (similar to Rossi et al., 2006) found an early sustained negativity (starting right after their baseline interval at 100 ms) in *untrained* subjects listening to Japanese ('Nihongo') sentences:

> *A broadly distributed early negativity (100–300 ms) in response to the word category violation occurred in the Japanese, untrained, and trained participants, but only Japanese natives displayed an anterior accentuation. [...] Whereas the frontally enhanced negativity in Japanese participants was expected based on results of previous experiments studying word category violation conditions (*Friederici et al., 1993*;* Neville et al., 1991*), the broadly distributed and early negativity observed in nonnative participants even prior to training came as a surprise. As knowledge of the phrase structure rules of Mini-Nihongo was certainly not present in participants before training, the broadly distributed negativity in untrained participants can only be related to processing of prosodic or phonological differences between the conditions* [p. 1238].

It should also be noted that those studies explicitly addressing prosody effects in PS violation paradigms (Eckstein & Friederici, 2005, 2006) have focused on other sentence types and did not investigate either pre-target effects or splicing effects at target word *onset*. They are, therefore, only of limited use to elucidate potential confounds in the German paradigm under discussion.

2.2.7. Local ELAN-like negativities without P600s

Above we argued that local ELAN effects (100–300 ms) in auditory studies can be typically accounted for by P600s superimposing sustained negativities. Does this mean that local ELANs, if they occur in absence of a P600, must be viewed as true indications of PS violations? We suspect that even those local effects, when elicited using the German paradigm, are more likely to be artifacts than real reflections of PS violations. Recall that in Friederici et al.'s 'suffix' study on delayed PS violations (2004) an undiscussed ELAN-like negativity between 100 and 300 ms can be seen before the word category information becomes available, in both the syntactic and the double violation. Interestingly, aligning and comparing this obvious artifact in the 2004 study with Hahne and Friederici's (1999) 'real' ELAN reveals surprising similarities: in both cases, the deviance occurred at left-anterior sites and with virtually identical onset latency and duration (see Fig. 4A–C). This similarity seems to suggest that not the (ungrammatical) target word but rather the (grammatical) pre-target prepositions may elicit effects that manifest as ELAN-like spillover effects on the next word, no matter whether the target word signals a PS violation or not. If so, one would expect to see a similar negativity also for control or filler conditions in which the preposition is followed by a grammatical noun. Whereas the 2004 study did not include any plots illustrating their filler conditions, the other German 'suffix' study by Friederici et al. (1996) did. They included a control condition that compared grammatical nouns within a PP (''zur Ver-edelung''/''for-the refinement'') with the same nouns in an ''anomalous'' condition ''wurde Ver-edelung''/''was refinement''). As hypothesized above, here the correct control (preposition + noun) seemed to elicit an 'ELAN'-like negativity in the same time window (Fig. 4D), suggesting that a certain word category or sound pattern in the context (rather than a word category violation on the target word) may be responsible for 'local' early left-anterior negativities of this kind.

2.3. Semantic blocking as evidence for 'syntax-first' models

As shown above, evidence supporting the ELAN component as a reliable marker of PS violations in phase-1 appears fragile at best. However, even if *all* evidence for ELANs turned out to be invalid, there is still another, largely independent, source of support that suggests a primacy of WC-based syntactic processing over semantic integration: **PS violations can block N400s** (and LANs). On a serial, syntax-first approach, the successful generation of a licit structural representation is a necessary pre-condition for either lexical/conceptual semantic, argument structure, or morpho-syntactic processes to apply.

Interestingly, unlike ELAN effects, evidence for ''blocking effects'' is actually fairly consistent across reading and auditory studies. In Friederici et al. (1999) double violations were tested consisting of both phrase structure and conceptual semantic violations (see example (9d)). These double-violation stimuli elicited the same ELAN/P600 pattern as the 'pure' PS-violations (9b), whereas the N400 observed in the pure *semantic* violation condition (9c) was entirely absent (i.e., blocked). This pattern has been replicated in a number of auditory studies (Hahne & Friederici, 2002; Friederici et al., 2004; Ye et al., 2006).[26]

(9)	a.	Correct:	*Das Haus wurde bald* ***gebaut***
			The house was soon built
	b.	PS viol:	*Das Haus wurde vom* ****gebaut***
			The house was by-the built
	c.	SemViol:	*Der Priester wurde bald* #***gebaut***
			The priest was soon built
	d.	Double Viol:	*Der Priester wurde vom* # ****gebaut***
			The priest was by-the built

A similar pattern of ''blocking'' has also been shown for double violations crossing word category and argument structure (Frisch et al., 2004; blocking N400 effects), and for word category and morpho-syntax (Rossi et al., 2005, 2006; blocking LAN effects).

It is important to point out that such blocking effects have been observed even in reading studies that did not show clear evidence of an ELAN component (Friederici et al., 1999; Frisch et al., 2004). Also, ERP blocking effects cannot be easily explained in terms of the kinds of baseline problems discussed above. In our view, blocking effects arguably provide the strongest empirical evidence in favor of syntax-first models and – as explicitly intended by Friederici (2002) – *against* the competing framework of interactive 'constraint-satisfaction' models that deny the primacy of syntax (e.g.,

[26] An anonymous reviewer points to the data pattern shown in Neville et al. (1991) as problematic for the hypothesis that ELAN and LAN effects should not co-occur. Recall that study reported two negativities: (i) the N125 (ELAN), and (ii) a later left temporal negativity between 300 and 500 ms. However, the blocking of N400s and LANs was suggested only for double violations. As 100–300 ms ELANs are often followed by sustained negativities even in studies of Friederici's group, we do not think Neville et al.'s (1991) data are *incompatible* with the model's claims (though this is not to say that the model has any ready account of such sequences of effects either). The same holds for Yamada & Neville's (2007) English/Jabberwocky differences.

Marslen-Wilson & Tyler, 1980; McClelland, St. John, & Taraban, 1989).

However, there are also a number of problematic findings. First, using the same paradigm as in (9) but with a semantic judgment task, Hahne & Friederici (2002, Experiment 2) found an N400 instead of a P600 effect for the double violation condition, suggesting that task effects modulate the actual ERP pattern (note that any account of this fact would run afoul the alleged immunity of phase-1 processes to strategic influences – see Section 2.1.3). Second, as mentioned above, Gunter & Friederici (1999) reported an N400 using the German paradigm but only with the preposition *vom* (by-the), which the authors argued involves the assignment of a thematic role. Explaining the fact that an N400, and not an ELAN was elicited in this way seems to imply that theta-role assignment (i.e., argument structure processing) can override and block phrase structure violations. As already noted, this is clearly inconsistent with the model.[27]

Third, van den Brink & Hagoort (2004) studied word category violations in a Dutch auditory experiment, again by replacing expected nouns (e.g., (10a)) with verbs (e.g., (10b)), however, with the prior context held constant (i.e., employing the target-manipulation design).

(10) a. *Het vrouwtje veegde de vloer met een oude **<u>bezem</u>** gemaakt van twijgen*
The woman swept the floor with an old <u>broom</u> made of twigs

b. *Het vrouwtje veegde de vloer met een oude ***<u>klieder-de</u>** gemaakt van twijgen*
The woman swept the floor with an old <u>mess-ed</u> made of twigs

Note that the disambiguating WC information in the PS violation condition (10b) was provided by the suffix *–de*. In addition, all target nouns were highly predictable semantically (i.e., high Cloze probability), such that the target verbs in the violation condition were expected to violate these semantic expectations in addition to the phrase structure. The authors reported an N400 and a sustained (L)AN, suggesting that semantic processing was apparently *not* blocked by the presence of a word category violation (for related findings in reading, see Zhang, Yu, & Boland, 2010). Furthermore, the onset of the N400 actually preceded the anterior negativity taken to reflect syntactic processes (see also van Petten et al., 1999, for a similarly early N400). Finally, once time-locked to the disambiguating suffix carrying the WC information, the LAN also showed a rather short latency of less than 300 ms (compatible with Friederici's predictions regarding ELAN 'latency shifts'). However, the lack of a blocking effect can be viewed as evidence against her syntax-first model.

How do the various blocking data inform the debate between strong syntax-first accounts and constraint-satisfaction/interactionist approaches? We suggest that studies which aimed to introduce WC violations via *prefixes* did not have the power to provide strong evidence relevant to distinguishing these views (even if such studies *had* succeeded in constructing "outright" violations – which they did not – see Section 2.1.2). This is because it seems reasonable to expect that an interactionist approach could model the relevant sorts of blocking effects for situations where WC information is encountered *before* lexical/conceptual semantic information becomes available (so long as syntactic information is allowed into the mix of competing factors in sentence processing). Conversely, as we will now discuss, studies which introduced WC information via *suffixes* and did not replicate the blocking of N400 effects (e.g., see (10a/b) above) have in fact been argued to be explainable in terms consistent with Friederici's approach. Therefore, a detailed understanding of the relevant studies and their findings is crucial.

In the following we will first describe in more detail the similarities and differences between two 'suffix' studies (Section 2.3.1), and then explain how the Dutch data can be convincingly described as compatible with Friederici's blocking hypothesis (Section 2.3.2). Then we will suggest a radically different account that has nothing to do with blocking whatsoever (Section 2.3.3).

2.3.1. Friederici et al. (2004)/van den Brink & Hagoort (2004)

Consider again (10b) from van den Brink and Hagoort (repeated here), and the examples in (11a–c) from Friederici et al.'s (2004) study:

(10)	b.	*Het vrouwtje veegde de vloer met een oude **<u>klieder-*de</u>***	**N400 → (E)LAN**
		The woman swept the floor with an old <u>mess-*ed</u> . . .	(***<u>no blocking</u>***)
		[The woman]$_{NP}$ [swept [the floor][with [an old → <u>noun</u>]$_{NP}$]$_{PP}$]$_{VP}$ . . .	
(11)	a.	*Der Strauch wurde **<u>verpflanz-t</u>*** . . .	**correct control**
		The bush was <u>replanted</u> . . .	
		[The bush]$_{NP}$ [was [→ <u>verb</u>]$_{VP}$]$_{IP}$	
	b.	*Das Buch wurde ***<u>verpflanz-t</u>**. . .*	**N400**
		The book was <u>replanted</u>. . .	
		[The book]$_{NP}$ [was [→ <u>verb</u>]$_{VP}$]$_{IP}$	
	c.	*Das Buch wurde trotz **<u>verpflanz-*t</u>**. . .?*	**<u>no N400</u>/ELAN**
		The book was despite <u>replanted</u>. . .	(***<u>'blocking'!</u>***)
		[The book]$_{NP}$ [was [despite [→ <u>noun</u>]$_{NP}$]$_{PP}$ verb]$_{IP}$	
	d.	*Das Buch wurde trotz **<u>verpflanzter</u>** Blumenbeete schliesslich vom Gaertner entdeckt.*	
		The book was despite <u>replanted</u> flower beds finally by the gardener discovered.	
		("Despite the replanted flower beds, the book was finally discovered by the gardener.")	

Both (10b) and the German double violation in (11c) replace an expected noun (within a PP) with a past participle. In both cases, the WC information is encoded on the suffix which elicited a (delayed) ELAN, but only (10b) yielded an N400 preceding this ELAN. Sentence (10b) differs from (11c) in that:

[27] A reviewer pointed out the possibility that the lack of the ELAN and elicitation of the N400 may be due to strategic processing as the violation condition could be identified based on a specific visual pattern (letter case). Besides the fact that this would violate the 'strategy immunity assumption' of Friederici's model, we argue that (a) strategic focus on physical features is more likely to suppress than to elicit N400s (e.g., Chwilla, Brown, & Hagoort, 1995) whereas (b) strategies are necessary to elicit rather than to suppress the ELAN in studies using the German paradigm.

(A) although both target words are initially WC ambiguous, the stem of the target element in (10b) (*'klieder'*/'mess') can stand alone as an **independent word** whereas the German stem (*verpflanz-*) in (11c) "has no independent word status" (i.e., is a bound morpheme, see Friederici & Kotz, 2003:pS11),
(B) the Dutch suffix ('*-de*') that renders the word stem into a verb is a **separate syllable**, whereas the German suffix encoding the WC information for a verb is part of the same syllable,
(C) the Dutch control condition used the **same context** and replaced the target word with an appropriate noun (i.e., *broom*, see (10a) above) whereas the German control in (11a) changed the context (i.e., the target word occurred within an Inflectional Phrase (IP) where a verb was actually expected),
(D) the Dutch sentence provided **more constraining syntactic and semantic context** both within the current PP/NP (i.e., the adjective *'oude'*/'old') and prior to the PP, such that,
(E) the Dutch target word could serve as the ***clause-final head noun*** whereas the German violation condition provided little context and required at least a head noun and a verb to complete the clause grammatically.

With these differences (A)–(E) on the table we turn now to possible explanations for the absence versus presence of N400 blocking for (10b) and (11c).

2.3.2. Explanations

Syntax-first views of blocking have at least two options to specify the temporal dynamics of semantic integration relative to WC information:

(i) <u>Bottom-up initiation:</u> always wait until WC information is *unambiguously* confirmed by the input and licit phrase-structural relations of the target word legitimize its semantic integration, or,
(ii) <u>Bottom-up blocking:</u> initiate semantic integration as soon as possible based on ***WC expectations*** (top-down predictions) but stop as soon as the input unambiguously signals a WC violation.
As neither of these principles can distinguish between (10b) and (11c), Friederici seems to suggest a third possibility:
(iii) <u>Bottom-up initiation and blocking</u>: *Wait until the input reveals* ***compatible WC information*** <u>(bottom-up initiation)</u> and continue unless or until the input signals a WC violation (<u>bottom-up blocking</u>) (Friederici & Kotz, 2003; pp. S11–S13).

In (iii), if mismatching WC information is available prior to or *concurrently* with semantic information, an ELAN is elicited and prevents ('blocks') *any* semantic integration. In other words, although top-down expectations are necessary to determine both WC compatibility *and* WC violations, the parser's behavior also *fully* depends on bottom-up information.

Dissimilarity (A) above – i.e., whether the relevant word stem has "independent word status" or not – is claimed by Friederici and Kotz (2003) to explain the distinct ERP patterns in (10b) and (11c). The suggestion seems to (have to) be that *whether the stem has independent word status matters for whether semantic integration is initiated for incoming WC ambiguous item* ***prior to*** *encounters with any WC disambiguating suffix.*[28] Again, both (10b) and (11c) are WC ambiguous until the crucial suffix material is encountered, but only in (10b) does the stem have independent word status (crucially: as a noun). Thus the idea seems to be that, despite the fact that downstream morphological information may indicate it is, in fact, a verb (as it does in (10b)), the independent word status of *klieder* ('mess') licenses syntactic integration, which then permits (phase-2) semantic integration processes to engage (yielding the N400). Then, when the suffix is encountered, the syntactic clash is detected, yielding the subsequent (E)LAN effect. In (11c), the idea is that the stem (*verpflanz-*) does not have the status of an independent word tied to a particular syntactic category, the syntactic integration required to engage semantic processes does not happen until the suffix is available, which however immediately signals a WC violation. Thus semantic integration is blocked, and no N400 effect is observed.

Using a strongly related (but logically distinct) argument that refers to **dissimilarity (B)** above and to "temporal integration windows" (Poeppel, 2003), Bornkessel-Schlesewky & Schlesewsky (2009) suggest that semantic integration can be blocked in (11c) – but not (10b) – because the WC violation is signaled *early enough*, i.e., on the *same syllable*. Both proposals seem, indeed, to account for the apparently inconsistent data.

In our view, however, there are problems as well. First, as discussed in Section 2.1.2., the past participle *'verpflanz-t'* in (11) can be used as an adjective and is, therefore, not ungrammatical even when the suffix is encountered. Second, Friederici's proposal seems to falsely predict an ELAN for sentences such as "*She loved her teach-er*" because the word stem "*teach*" temporarily signals the ungrammatical WC of a 'verb' (and thus a WC violation). Moreover, since the 'bottom-up initiation' principle would falsely predict delayed N400s for polymorphemic words (i.e., semantic integration can start only after the entire word is available), only the 'bottom-up blocking' principle seems plausible. This means that semantic integration is (most likely) initially based on ***top-down*** *expectations* (including syntactic WC projections) and ***bottom-up*** *semantic* information (but not bottom-up WC information). Friederici's distinction between stem morphemes that do versus do not have "independent word status" does not play a role unless it can be shown that bound word stems prevent retrieval of lexical-semantic information. Note, however, that Bornkessel-Schlesewsky and Schlesewsky's 'temporal integration' proposal may be more compatible with this principle and could replace Friederici's original architecture.

The scenario outlined above re-interprets what *'syntax-first'* might mean with respect to semantic integration: initial semantic processing is largely determined by *context-driven top-down* expectations (initial parsing stage), not by WC information retrieved (bottom-up) from the target input (see our Fig. 1). Therefore, whereas target-manipulating paradigms keep the initial parsing stage constant across conditions, effects related to semantic processing (e.g., blocking) in *context-manipulating* PS violation paradigms may primarily depend on contextual differences rather than PS violations. Intriguingly, to date blocking effects have been observed *only* in studies using the context-manipulating German paradigm.

2.3.3. An alternative

Our proposed explanation of the ERP differences between the two studies is independent of PS violations and blocking effects. Taking the above mentioned **dissimilarities C, D** and **E** seriously, we suggest that the distinct syntactic and semantic contexts (both between (11a), (11b) and (11c) within the German study and between (10b) and (11c) across studies) are key to understanding the ERPs. Within the German study, both (11a) and (11b) predict a verb that functions as a predicate and has to be semantically integrated with the *preceding* subject NP (indicated by arrows), thus eliciting the N400 in (11b). By contrast, in (11c), the preposition

[28] About a sentence such as *"He knows how to *investment"*, Friederici & Kotz (2003, p. S12) write: "if the word stem itself represents an independent lexical entry carrying word category information (e.g., *invest-ment*, first part is a verb)" then "the incrementally working parser would, once word category information of the stem (*invest* – verb) is available, try to integrate this part into the preceding context and detect a syntactic word category mismatch only later."

('trotz'/'despite') projects an NP (within a PP) predicting a head noun (but multiple other WCs would be grammatical as well, including: adjectives, adverbs, determiners, prepositions and conjunctions). It is quite conceivable that any incoming word (or word beginning) would be preferably interpreted as one of these permitted word categories, potentially prioritizing those that require the least complex projections (e.g., an adjective modifying the already projected head noun is more likely than another preposition requiring the projection of several new phrases). According to the scenario above, the word stem of the incoming word (verpflanz-) will be interpreted based on these expectations, i.e., most likely as an adjective. However, as this adjective is supposed to modify the *subsequent* head noun (which is not available yet), semantic integration has to remain very limited. Note that a semantically quite plausible continuation is still possible (as illustrated in (11d)), such that the weak prior context (the subject NP 'the book') does not suggest a strong semantic mismatch – and thus no N400. One could argue that this scenario is not very different from Friederici's ideas of how syntactic structure determines semantic integration. However, if our scenario is right, even the perfectly grammatical adjective in (11d) is not supposed to elicit an N400, whereas it should in Friederici's framework (otherwise the entire idea of blocking would not make sense).

Employing sentences such as (11d), these competing hypotheses are, of course, easy to test. Note also that in both (10b) and (11b), the contextual constraints *are* sufficient to secure a semantic integration of the incoming word stem such that a conceptual-semantic anomaly results (thus explaining the *presence* of an N400). Importantly, given that the main decisions as to how the bottom-up semantic information will be integrated depend primarily on the initial *syntactic* parsing stage and its projections, our scenario can still be understood as a 'syntax-first' account. In other words, we do think that projected hierarchical phrase structural relationships strongly constrain how incoming words *could* be integrated even if they are implausible (e.g., *'man bites dog'*).[29]

A final comment concerns the impact of the *specific selection* of prepositions (and other categories). We predict that 'every single word matters' and modifies the specific nature of the parser's expectations. Imagine an experiment that uses only one preposition, such as *'vom'* (by-the), to introduce the agent of passive sentences. As agents are prototypically animate noun phrases (either just in our hypothetical experiment, or even more generally, e.g.. Bornkessel-Schlesewky & Schlesewsky, 2006), an adjective following the preposition *'vom'* would be predicted to modify the projected (animate) noun. However, if this adjective turns out to be an implausible modifier of an *animate* agent (say 'brushed', which of course works well with *inanimate* nouns as in *'the brushed steel'*), this conceptual-semantic incompatibility with the current *projection* may already elicit an N400, even *in absence* of the head noun. We suspect that phenomena like this may be what elicited the ERP pattern of Gunter & Friederici (1999), where German PS violations such as "*Die Zähne wurden vom gebürstet*" ("*The teeth were by-the brushed*") did actually yield an N400. But no ELAN.

3. New landscape

Our critical analysis of the ELAN literature was divided into three parts. The first part (Section 2.1) argued that evidence for the modality independence of ELANs is weak, that ELANs are clearly not specific to "outright violations", and that ELANs are influenced by (or even reliant on) top-down/strategic processes in studies using the German paradigm. The second part (Section 2.2) argued that there is good reason to think that a great deal of the existing data in the ELAN literature is ambiguous: such effects may either be legitimate markers of early syntactic processing *or* may be irrelevant contextually-driven artifacts. Finally (in Section 2.3) we examined the logic of so-called "blocking effects", arguing that a more careful look at these findings reveals that they are actually far less clear as evidence favoring syntax-first approaches than may be apparent at first blush. We also offered an alternative (and easily testable) account of the relevant empirical data which eschews reference to blocking completely.

In what follows we provide a survey of the new landscape of empirical generalizations which emerges if we take the combined weight of our main points seriously. We discuss our view of ELAN effects in reading (Section 3.1) and in auditory processing (Section 3.2), before closing (in Section 3.3) with some remarks concerning consequences for other areas of investigation (e.g., studies of patient groups or late acquired second languages (L2)).

3.1. PS violations in reading

First, it seems necessary to distinguish between the short-lived (and inconsistently reported) local ELAN effects in the visual domain versus more sustained negative deflections which – we have argued – may emerge as the general case in the auditory domain. In reading, the few studies that found ELAN-type effects were those which have used asymmetric violation paradigms (Neville et al., 1991; Friederici et al., 1999; Yamada & Neville, 2007). These studies, as we have argued, run the risk that context (spillover) effects may have been misinterpreted as early target word effects (though whether this is *generally* the case, or not, remains to be demonstrated).

Moreover, even if these very early effects (e.g., Neville et al.'s N125) in reading are real, they may not be best understood as reflecting the action of an *amodal* syntactic module. MEG data (Dikker et al., 2009) suggest that brain responses (M100) in the latency range of the N100 component in ERPs may index (syntax-driven) patterns of perceptual mismatch effects generated in the visual cortex (see fn.6), which may account for short-lived effects such as Neville et al.'s (1991) N125.[30] A further important feature of the Dikker et al. (2009) data is that this early M100 effect did not distinguish between violations realized on bound word-final suffixes and free closed-class words such as prepositions (as in the (Neville et al., 1991) materials – see (1)). This finding, incidentally,

[29] In fact, if semantic integration follows hierarchical syntactic relationships (e.g., first within an NP, then within an IP, etc.), it might be possible to track hierarchical syntactic integration processes by studying syntactically and conceptually complex sentences with N400 components. The amplitude gradient of N400s along with the high temporal resolution of ERPs would be an ideal tool to investigate both the temporal order of integration sub-processes and, e.g., the impact of depth of syntactic embedding on semantic priming in sentence contexts.

[30] (1) With respect to reading studies, short-lived perceptual effects modulating the N100 would also entirely explain the surprising finding that the elicitation of ELANs was dependent on strong visual contrasts (Gunter & Friederici, 1999). Their Fig. 1 suggests larger onset and offset components (N100s and P200s) for (a) longer words (both pre-target and target) and (b) stronger visual contrast. That is, the "ELAN" (at F3) was most likely a word-*onset*-N100, enhanced both by the visual contrast itself and by a DC-offset artifact (due to different pre-target offset components in the baseline interval). Neither the correspondingly large target word-*offset* components (e.g., an enhanced target-word offset-P200 in the violation condition, also at F3, around 500 ms, i.e., 200 ms after word offset) nor the obvious pre-target word-offset differences between high and low contrast (−200 to 0 ms) illustrating these relationships were addressed by the authors. (2) Similar phenomena may, of course, also hold for early effects in the auditory domain. In fact, van den Brink Brown, and Hagoort (2001), van den Brink and Hagoort (2004) have discussed N200-like effects along these lines, and Connolly and Phillips (1994) reported a phonological mismatch negativity (PMN) for words whose initial phonemes that mismatched with those of the expected target word (although these components emerged rather late, around 350 ms). Such effects could in principle also contribute to auditory ERP effects in PS violation studies, especially when predictable word category markers are involved. However, the available evidence suggests relatively short-lived N200 and PMN effects that do not seem to account for what we view as the general pattern for auditory PS violations, i.e., sustained negativities (and P600s).

undermines a reinterpretation of reading data from Hagoort et al. (2003) offered by Friederici & Weissenborn (2007). Hagoort et al. reported a PS violation effect (a bilateral anterior negativity; "AN") between 300 and 500 ms, but no ELAN. Friederici and Weissenborn argue that this pattern has to do with the late availability of word category information, which was encoded on a suffix in Hagoort et al.'s materials. But the Dikker et al. MEG data show that, unlike in auditory studies, the *position* of word category information in reading studies (prefix, word stem, suffix) may not affect the latency of early effects.

Those reading studies which have avoided the context-manipulation approach have uniformly failed to replicate ELAN-type effects (see Appendix, Table A1), while LAN-like effects in later latency ranges (~300–500 ms) appear to be the most prevalent pattern. This suggests that there are, in fact, no systematic differences between the timing of phrase-structure versus morpho-syntactic (or semantic) violation effects in reading. This is explicitly shown in a study by Martin-Loeches et al. (2005), where a target-manipulation approach was used. If correct, the lack of latency differences would actually be consistent with data from auditory studies that have similarly avoided potential context effects (e.g., Hasting & Kotz, 2008), and cast severe doubt on the serial nature of phase-1 and phase-2 in Friederici's model. In sum, visual PS violation effects, like other syntactic violations, seem to usually elicit LANs between 300 and 500 ms, whereas early local ELAN-like findings are rather rare and remain ambiguous until further research clarifies their status.

3.2. PS violations in speech

In auditory studies, by contrast, reports of early local effects seem, we suggest, to be typically due to *sustained negativities* that are superimposed by large P600 components. In all PS violation studies we are aware of that show either strongly reduced P600s or no P600 at all, sustained anterior negativities rather than local ELANs have been observed (e.g., Friederici et al., 1993; Hahne & Friederici, 1999; Exp. 2; Hahne & Friederici, 2002; Exp. 2; Hasting & Kotz, 2008; Exp. 2; Ye et al., 2006; Rossi et al., 2006, especially low proficiency groups; Pakulak & Neville, 2010, low proficiency L1 group). Conversely, all auditory studies that report *local* ELAN-like effects, also show large P600-like positivities (sometimes preceded by frontal P3a-like components; e.g., Hahne and Jescheniak, 2001), whose respective onset latency generally appears to account for the temporary attenuation/cancellation of ongoing frontal negativities. A common pattern across studies seems to be that the sustained negativity remains largely unaffected at lateral anterior sites (F7, F8), whereas it is increasingly affected (attenuated) the closer the electrode site is to PZ/P4 (where the P600 tends to reach its maximum amplitude). As one would expect, in a considerable number of such studies the frontal negativity re-occurs as a 'late component' after the P600 has reached its peak amplitude (see our Fig. 4A for an illustration), especially at medial electrodes such as F3, Fz, F4, and Cz (e.g., Hahne & Friederici, 1999, Exp. 1; Pakulak & Neville, 2010, low proficiency group; Mueller et al., 2005, Japanese group; Rossi et al., 2006, most groups).[31]

Given this pattern, it is not surprising that fMRI data collected by Pakulak (2008) provided evidence *"that one of the neural generators of this component is anatomically close to the neural generator of the anterior negativity in the 100- to 300-ms time window, suggesting that the negativity over anterior sites in the later time window indexes processes similar to the negativity in the earlier time window"* (Pakulak & Neville, 2010). More surprising is the fact that (to our knowledge) virtually all studies that discuss the late part of this negativity at all, tend to treat it as an "effect" or "component" that is distinct from the early part.[32] It is not unlikely that this interpretation may be strongly influenced by Friederici's 3-phase model, according to which a 100–300 ms time window must be viewed as the 'gold standard' to test for ELAN components reflecting phase-1. In our view, such analyses should be complemented by direct comparisons between early and late effects that take the presence of a P600 and its respective latency, duration, amplitude, and scalp distribution into account. For example, if a bilateral negativity is partially superposed by a right-lateralized parietal P600, this may result in an apparent left-anterior local negativity.

3.2.1. Sustained negativities

Apart from its longer duration, the sustained negativity in the auditory modality differs from effects in reading studies in that (i) its onset latency does depend on the availability of relevant word category information in the speech stream, (ii) when time-locked to this information, its onset latency is short (within the first 100–250 ms) and (iii) these early effects can be found even in target-manipulating studies that avoid context effects (e.g., Hasting & Kotz, 2008; van den Brink & Hagoort, 2004). These three characteristics are clearly in line with Friederici's model. While sustained negativities with rather early onsets appear typical for PS violations, are they specific to this kind of violation?

First, we have shown that in at least some studies (e.g., Rossi et al., 2005, 2006) offset effects due to baseline problems must have resulted in similar sustained negativities, since this ERP pattern occurred *prior to* the word category information. In our view, lexical (word class) and acoustic/prosodic context differences between PS violation and control conditions were responsible for the Rossi et al. effects. This is largely in line with Mueller et al.'s (2005) 'prosodic' interpretation who (similar to Rossi et al.) also found sustained negativities with 'ELAN-like' onset latencies of 100 ms that simply *could not* be due to the PS violation because the untrained (German) subjects did not understand the language (Japanese). In addition to these obvious cases, the extremely short onset latency of the sustained negativity in Ye et al. (2006) between 50 and 100 ms also seems to point to a problem with their (−1100 to 0 ms) baseline, since this (significant) 'syntactic violation' effect appeared even prior to the typical latency of perceptual ERP effects generated in the primary auditory cortex (for related discussion see Lau et al., 2006 & Dikker et al., 2009). In sum, there seem to be at least four studies using context-manipulating PS violation paradigms in no less than four languages, whose sustained negativities cannot be explained by the PS violation and therefore must be artifacts (Rossi et al., 2005, 2006, in German and Italian;

[31] Note that this pattern holds also for Rossi et al. (2006), where we argued that (at least the early part of) the sustained negativity is likely to be due to an offset artifact. What this study also illustrates is that in those groups that show a delayed P600 (e.g., low proficiency L2 learners of German), both the attenuation of the negativity *and* its re-occurrence are delayed. A similar pattern was observed by Sabisch and colleagues (2006) in children who also show a somewhat delayed and long-lasting P600 (~600–1300 ms).

[32] A few studies (e.g., Friederici et al., 1993) demonstrate that the early part of the negativity (100–250 ms) has a significantly different scalp distribution than the later part (250–700 ms), which would indeed suggest distinct neural generators. However, this topographic difference is very likely due to the presence of a P600 in the violation condition during the later time window. Note that in Friederici et al. (1993) as well as a number of other studies (e.g., Rossi et al., 2006, low proficiency groups; Ye et al., 2006; Pakulak & Neville, 2010, low proficiency group), the P600 was so small compared to the amplitude of the sustained negativity that it was barely able to cancel out the negativity and display a positive net amplitude even at central electrodes (where most other studies/participants usually show large P600s, e.g., the high proficiency groups in Rossi et al. and Pakulak and Neville). In our view, the standard quantification of ERP components (in terms of net amplitudes measured *relative to the baseline*) are often insufficient to adequately characterize the superposition of concurrent ERP components. For example, in Ye et al. (2006), only the presence of a P600-like positivity can explain why a sustained negativity (50–1000 ms) at frontal lateral electrodes (F5, F6) resembles a more transient "N400-like" negativity (100–500 ms) at more posterior sites.

Mueller et al., 2005, in Japanese; Ye et al., 2006, in Chinese). Since all (other) auditory ELAN studies using the German (context-manipulating) paradigm are prone to similar artifacts and, additionally, have the problem of prefix ambiguities (*ge-/be-*), the status of their findings must be viewed as ambiguous.[33]

On the other hand, data from studies using target-manipulating or balanced designs (thereby avoiding context effects; e.g., Hasting & Kotz, 2008; van den Brink & Hagoort, 2004) clearly provide strong evidence that sustained negativities with onset latencies within 100–300 ms can also be *true* reflections of syntactic violations. However, Hasting and Kotz demonstrated similarly early effects for morpho-syntactic agreement violations when the ERP analysis is appropriately time-locked to the disambiguating morpheme. This conflicts with the prevalent view according to which morpho-syntactic violations typically elicit *later* (300–500 ms) LANs, corresponding to phase-2 in Friederici's model. In this context it is worth mentioning that the vast majority of morpho-syntactic violation data come from reading rather than auditory studies, e.g. Friederici's (2002) paper (presenting her model for "auditory sentence processing") refers *only* to reading studies to support LAN effects in the 300–500 ms time range. One of the few auditory ERP studies investigating morpho-syntactic effects is that by Mueller et al. (2005; for case violations in Japanese) that reports a bilateral "*negativity with an anterior preponderance peaking at about 400 ms post-stimulus onset*", which however seems to re-occur after the P600 and may, therefore, alternatively be characterized as a sustained anterior negativity superimposed by a P600. Moreover, since the latency of the negativity was measured from word onset, analyses time-locked to the relevant information would likely have resulted in onset latencies in the classical ELAN time window. A similar pattern can be seen in Rossi et al.'s (2006) agreement data (which do not have a baseline problem), especially in native speakers.[34]

In sum, it seems that the pattern of sustained negativities plus P600s constitutes the reliable pattern seen in connection with syntactic PS violations in auditory ERP studies, but is *not* specific to PS violations. In some (but not all) cases, the sustained negativities are simply artifacts. In other cases, they also manifest for morpho-syntactic violations in the auditory domain (with non-distinct onset latencies when time-locked to the relevant information).

3.2.2. Working memory and sustained negativities

If sustained negativities rather than local ELAN effects are characteristic for syntactic violations – what do they reflect? Working memory has frequently been appealed to as a concept of potential utility in bringing a variety of "LAN-type" effects seen in language processing under one umbrella. Sustained negative-going shifts over left/anterior scalp regions have been convincingly connected to working memory in a variety of language-related domains: (i) filler-gap dependencies (Fiebach, Schlesewsky, & Friederici, 2001; Fiebach, Schlesewsky, & Friederici, 2002; King & Kutas, 1995; Kluender & Kutas, 1993; Phillips, Kazanina, & Abada, 2005), (ii) semantic violations involving temporal relationships, mood, and modality (Dwivedi, Phillips, Lague-Beauvais, & Baum, 2006; Münte, Schiltz, & Kutas, 1998); (iii) subvocal rehearsal (Ruchkin, Johnson, Grafman, Canoune, & Ritter, 1992), and (iv) mapping relationships between logical semantics and discourse representation (Steinhauer, Drury et al., 2010; Drury et al., in preparation; van Berkum, Koornneef, Otten, & Nieuwland, 2007).

A parsimonious approach would be to understand sustained negative shifts in syntax studies to be a sub-species of this more general class of working memory related effects. Steinhauer et al. (2010) suggest that such negativities may appear whenever linguistic violations "*result in the need to maintain representations of incoming material in an unintegrated format, thus imposing an increase in demand on working memory resources. This notion would assimilate well with the finding of sustained negative deflections for filler-gap types of relationships [...] and with the general finding of such effects in connection with rehearsal of words*."

Under this perspective, the fact that these negativities may be more likely to occur with auditorily than visually presented word category violations is actually expected and may be tied to the assumption that only auditory information inevitably enters the phonological buffer of the working memory system. Moreover this view also predicts that variability in such effects should be influenced by a range of factors, including individual differences in working memory capacity, task requirements, and the complexity of sentence materials, in line with previous findings (King & Kutas, 1995; Martin-Loeches et al., 2005; Vos, Gunter, Schriefers, & Friederici, 2001a, Vos, Gunter, Kolk, & Mulder, 2001b).

We have already argued that these sustained negativities do not merely precede but rather largely overlap with P600-type effects. According to any serial approach linking the onset of these negativities to the initial disruption of PS generation, such effects *must* precede the more "controlled" processes reflected by P600s. In contrast, on a working memory related view, there is nothing inevitable about this ordering of negative/positive-going effects. That is, in principle these negativities may even follow the P600-type effects. If that were the case it would have consequences for how we understand what P600 effects are as well (i.e., viewing them as reflecting "reanalysis/repair" might be too narrow a characterization, see Bornkessel-Schlesewsky & Schlesewsky (2006, 2008), Friederici & Weissenborn (2007), Kuperberg (2007), & Steinhauer et al., 2010 for discussion of related ideas).

Interestingly, LAN-type effects following P600s have in fact been observed in studies examining logical semantic/pragmatic anomalies (referred to as a P600/L-LAN pattern – for "late LAN" – in Steinhauer et al. (2010); see also Drury & Steinhauer, 2009 for discussion). Further, late sustained negativities following N400 effects for lexical/conceptual semantic anomalies are also not unheard of (see Steinhauer et al. (2010) for discussion and references).[35] Whether

[33] In addition to offset effects resulting in sustained negativities, our recent auditory ERP study using a completely balanced design (similar to example (6) above) strongly suggest that context effects that in unbalanced designs may look like violation effects, can also occur during the first 300 ms (Pauker & Steinhauer, 2010). Artifacts of this kind are likely to account for relatively early short-lived differences, which were present in Friederici et al. (2004) for PS violations only, but prior to the disambiguating suffix providing word category information, and which were not discussed by the authors. Future research needs to clarify the origin of such early left anterior artifacts (even if they turn out to be statistically non-significant) as they share all features of 'classical' local ELAN components and can easily be mistaken for real ELAN effects, especially in cases where their status as an artifact is less obvious than in Friederici et al. (2004). Note, however, that the plausible (delayed) PS violation effect that *followed* the disambiguating suffix (and was discussed in that paper) is, again, in line with a sustained negativity followed by a P600, which we have argued is the general pattern for PS violations in auditory studies.

[34] In Rossi et al.'s (2006) proficient L2 groups, the P600 peak is somewhat delayed and has a longer duration, such that a re-occurring frontal negativity can be seen only in the last 100 ms prior to the end of their 1500 ms epoch. (Note, however, that high proficiency L2 learners of German show a sustained negativity between 400 and 1500 ms at the left-anterior electrode F7). Low proficiency groups do not show any such anterior negativities for agreement violation.

[35] The finding of sustained negativities following N400 effects in conceptual anomalies is, of course, not commonly reported. As discussed in Steinhauer et al. (2010) and in our main text discussion above, many factors may contribute to the presence/absence of such effects (task, individual working memory capacity differences, etc.). Further, in studies targeting the N400, later negative-going shifts may have been missed by analyses concentrating on early latency ranges. It should also be noted that the specifics of data preprocessing may critically impact the profile of these negativities. For example, applying a band-pass or high-pass filter to the data could (depending on the cutoff) eliminate sustained effects while preserving the initial (more phasic) shift into the negative (or positive) amplitude range, yielding the appearance of a local/transient effect. For example, Friederici, Steinhauer, and Pfeifer (2002) applied this strategy to MEG data on word category violations and interpreted the resulting early local effect as the MEG analogue of the "ELAN".

all of these negativities (both pre- and post-P600 onset) are indeed best understood as indexing a single mechanism requires further investigation.

3.3. ELANs in language acquisition and breakdown

Our final points concern the way that our understanding of the nature of "(E)LANs" impacts the interpretation of group differences reported in studies of first (L1) and second (L2) language acquisition, as well as in clinical investigations of neuropathology. In children up to an age of 10 years (Hahne, Eckstein, & Friederici, 2004) and in both late L2 learners (e.g., Hahne, 2001) and Broca's aphasics (Friederici et al., 1998), the absence of ELAN effects for word category violations has been argued within the framework of Friederici's model to reflect the inaccessibility of early/automatic (phase-I) syntactic parsing mechanisms. However, if interpretations of ELAN effects need to be reconsidered, our understanding of such group differences will consequently also be affected. In the following we will briefly illustrate this for the domain of L2 acquisition.

We have discussed elsewhere (Steinhauer et al., 2009) that L2 proficiency (rather than the age of L2 acquisition) may systematically influence the ERP profiles in second language learners. There, we argued that such proficiency-dependent changes may include both qualitative changes (i.e., the emergence of new components) and quantitative changes (e.g., the increase of P600 amplitudes). In connection to the specific case of PS violations, it is interesting that both Mueller et al. (2005, in second language learners; see also Rossi et al. (2006), for similar patterns in Italian) and Pakulak and Neville, 2010, for native speakers demonstrate that low proficiency groups show *sustained negativities and small or no P600s*, while high proficiency groups display rather *local negativities and large P600s*. Across groups, both studies interpreted the early portion of these negativities preceding the P600 as an (E)LAN-like component, while the later portion was viewed as a potentially distinct (and 'unexpected') component (Pakulak & Neville, 2010). They interpret these findings as suggesting that, "*early and immediate sentence parsing processes may operate differently, or be used differently, in adult monolingual native speakers who differ in their linguistic proficiency when processing their native language*" (p. 2738).

However, in line with our discussion above, visual inspection of the ERP plots in both studies strongly suggests that a sustained negativity was actually present in *all* groups. Thus, the entire pattern of ERP differences between proficiency groups can potentially be explained by the presence of an enhanced P600 for high proficiency which superimposed the sustained negativity to a larger extent compared to low proficiency. The early onset of the posterior P600 observed by Pakulak and Neville (2010) (apparently present around 0 ms at right-posterior sites) may also have contributed to the different scalp distributions in the early (100–300 ms) time window. That is, in the high-proficiency group, the sustained negativity may have 'survived' only at those (left-anterior) electrodes that were maximally distant from the P4 electrode showing the largest P600 amplitude (cf. their Fig. 2).

In other words, the main differences between the high/low proficiency groups may actually be reflecting differences in the P600 which, due to the temporary superposition/cancelation of effects, also results in the appearance of very different looking anterior negativities. [36]

We believe it is a mistake to consider group differences with respect to the profile of anterior negativities independently of co-occurring P600-type effects. Also, recall that most (if not all) of these group studies were conducted using unbalanced paradigms in the auditory domain which, as we have argued above, are especially prone to ambiguities between "real" violation effects and context driven artifacts (e.g., spillover and offset effects). Moreover, "real" violation effects in studies that used the German paradigm may actually have crucially depended upon participants' abilities to develop processing strategies (e.g., so that category ambiguous prefixes like *ge-/be-* could be systematically interpreted as *verb specific*). Thus, group differences in the elicitation of corresponding ERP components could be alternatively re-interpreted in either of two ways: (i) as differences related to the various factors which can yield context-driven artifacts (e.g., processing of pre-target word-class distinctions, prosody, etc.) or (ii) as differences in the abilities to acquire strategies (or implicit/statistical knowledge about linguistic elements) over the course of an experiment. While evaluating these alternative interpretations of group differences corresponds to a straightforward agenda for future research, assuming that children, late second language learners, and Broca's aphasics may indeed have such problems seems to us to be not entirely implausible.[37]

Note, none of this is to say that these groups do not differ from healthy, highly proficient, monolingual adults. What is unclear, in our view, is precisely *how* they differ, and whether previous work claiming a specifically syntactic locus for these differences has succeeded in demonstrating this.

4. Conclusion: What does all of this leave us with?

If even some of our critical points raised here turn out to be valid, then we are at risk of losing the elegance, scope, and clarity of predictions provided by Friederici's model. This begs the obvious question of what could replace it. Obviously, it is highly unsatisfactory to just vaguely align N400s with (e.g.) "lexical access/retrieval", P600s with "integration" and sustained negativities with "working memory". Even though this may be true, it is embarrassingly uninformative. But still, we believe this may be progress compared to an elegant but incorrect account. A precondition for evaluating alternative frameworks is to have a clear picture of the relevant empirical generalizations that any explanatory framework has to account for. The objective of the present paper was to contribute to charting out this territory.

Our present view is that the three-stage architecture of Friederici's model may have to be modified as there seems to be little evidence for a first phase exclusively dedicated to phrase structure processing. Moreover, context-driven top-down processing may play a larger role than assumed by the current version of this model (see Section 2.3). Future ERP studies on syntactic processing will be critical in specifying the details of a revised model. As various methodological problems in previous work appear to have resulted in misleading or ambiguous data, we believe that a discussion about minimal methodological requirements for psycholinguistic ERP experiments may be in order, in terms of both (a) experimental designs and paradigms and (b) data

[36] With respect to developmental ERP differences between children (up to an age of 10 years) and adults, the larger (and earlier) P600 superimposing the sustained negativity in adults seems to account for these patterns in a similar way as in second language acquisition, whereas the shorter onset latencies of frontal negativities with increasing age (Hahne et al., 2004) appear to be a distinct effect.

[37] Consider some examples. Context effects, we have argued, may arise in virtue of word class (e.g., open vs. closed) differences on pre-target elements. Thus, group differences in *these* effects would be expected to yield different spillover/offset artifacts on the target words. That such group differences actually exist has been shown, for example, for Broca's aphasics, who do not show N280/N400-700 effects for closed-class (relative to open-class) words (Brown et al., 1999). Group differences in prosodic processing could also be similarly implicated (given that in auditory studies prosodic factors also contribute to context effects; Mueller et al., 2005).

analysis and data presentation. We hope the present paper will assist in developing heightened awareness of these issues and, in the long term, better standards in the field.

Acknowledgments

This work was supported by grants awarded to Dr. Steinhauer by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; Project # 201876), by the Natural Sciences and Engineering Research Council (NSERC; # RGPGP 312835-2005), by the Social Science and Humanities Research Council (SSHRC; #410-2007-1501), and by the Canadian Institutes of Health Research (CIHR; # MOP -74575).

Appendix A

The following two tables provide a representative summary of ELAN-related ERP studies, both for reading (Table A1) and auditory presentation (Table A2). In these summaries, we have restricted our attention to experiments that have included phrase structure / word category violations of the sort we focus on in our critical discussion. Where these studies included other types of violations that are relevant (e.g., in studies examining "blocking" effects), example stimuli and ERP findings for those are included as well, as are other relevant manipulations such as task (e.g., acceptability versus sensicality judgment, as in study [5] in Table A2), properties of the visual display (e.g., high versus low contrast in study [8] in Table A1), or special contextual manipulations specific to particular studies (e.g., the +/-ellipsis manipulation in study [16], Table A1).

Tables A1 and A2 adopt the following conventions. First, studies which employed TARGET-manipulation paradigms, mixed TARGET/CONTEXT-manipulation designs, or CONTEXT-manipulation with appropriate additional controls are shaded in grey – all others (unshaded) rows correspond to studies using only CONTEXT-manipulation.

In the example stimuli representing the experimental contrast(s) of interest, target words to which ERPs were time-locked are indicated in **bold**, and the pre-target words are <u>underlined</u>. Target words in correct/grammatical control conditions are indicated by a ✓,and violating words are indicated by * (double violations are marked by **). ERP findings are summarized relative to the three phases (I-III) postulated in Friederici's (2002) model. In cases where effects do not clearly fall into a single hypothesized latency range corresponding to a particular phase, the relevant phase columns are collapsed (and latency of effects are provided throughout where relevant). (E)LAN type effects are referred to descriptively as follows:

- LAN = left anterior negativity
  - if this occurs in Phase I latency ranges, this can be regarded as an "ELAN" finding
  - we reserve this designation for effects that have been demonstrated to be both left and anterior in scalp distribution
- Where LAN-like effects have not shown either left lateralization or clearly anterior scalp distributions, the following intuitive abbreviations are used:
  - AN = anterior negativity (no lateralization of effect)
  - LN = left lateralized negativity
  - (LA)N, (A)N = where descriptors appear in parentheses, these refer to descriptive generalizations regarding the perceived (but not statistically demonstrated) distributional properties. Thus, (A)N is a broadly distributed negativity with an anterior maximum, (LA)N is a broadly distributed negativity with a left anterior maximum, etc.
  - PN = posterior negativity (this label is used in a case where the observed effect might be categorized as belonging to the family of N400-type effects, but where the authors did not interpret the finding in this way).
- **???** = this appears in the table where the presence/absence of early ELAN-like effects is impossible to determine due to obvious baseline confounds (e.g., where large effects at or before target word onset are clearly present).

Labels for conditions. In addition to phrase structure (PS) violations, some of the studies included tested lexical/conceptual semantic (SEM), argument structure (AS), and morpho-syntactic (MS) violations. Where violation types were combined to create "double" violations, this is indicated with the "+" symbol (e.g., PS + SEM = a combined phrase structure and conceptual semantic violation). Finally, note that the studies are indexed with letters in square brackets [a–s] in the leftmost column of Tables A1 and A2 indicating the sections of our main text discussion where these studies are mentioned. An outline of the present paper follows these Tables where the correspondence between these letter indices and sections is indicated.

Table A1
Reading Studies.

#	Study [main text]	Year	Example stimuli	CONDITIONS	Phase I (100-300)	Phase II (300-500)	Phase III (> 600)
[1]	**Neville et al.** *[b-d,i,m,o]*	1991	The scientists criticized Max's proof ✓**of** the theorem				
			The scientists criticized Max's ***of** proof the theorem		**LAN**/N125	**LN**	**P600**
[2]	**Hagoort et al.** *[a,d]*	1993	De echtgenoot schrikt van de nogal emotionele ✓**reactive** van zijn vrouw *The husband [is startled] by the rather emotional response of his wife*				
			De echtgenoot schrikt van de emotionele nogal * **reactive** van zijn vrouw *The husband [is startled] by the emotional rather response of his wife*		—	—	**P600**
[3]	**Münte, Heinze, & Magnun**	1993	You ✓**spend** / My ✓**laughter**				
			Your * **write** / You * **administration**		—	**LAN**	—
[4]	**Friederici, Hahne, Mecklinger** (Exp.2) *[c,e,k]*	1996	Das Metall wurde ✓**veredelt** von dem Goldschmied den man… *The metal was refined by the goldsmith who was…*				
			Das Metall wurde zur * **veredelt** von dem Goldschmied den man… *The metal was for refined by the goldsmith who was…*		—	**LAN** 350-500	**Late LAN** 700-1100 + **P600** 500-1000
[5]	**Ainsworth-Darnell et al.** *[d]*	1998	Jill entrusted the recipe to ✓**friends** before she…				
			Jill entrusted the recipe * **friends** before she…		—	—	**P600**
[6]	**Friederici, Steinhauer, & Frisch** *[b,c,g,l,m,o]*	1999	Das Haus wurde bald ✓**gebaut** *The house was soon built*				
			Der Priester wurde bald #**gebaut** *The priest was soon built*	SEM		**N400**	
			Das Hausz wurde vom ***gebaut**. *The house was by the built*	PS	**LAN** (F7)	—	**P600**
			Der Priester wurde vom *#**gebaut** *The priest was by the built.*	PS+SEM	**LAN** (F7)	—	**P600**
[7]	**Gunter, Friederici & Hahne** *[c]*	1999	Die Tür wurde ✓**geschlossen** *The door was closed.*				
			Der Kanzler wurde am * **gewählt** *The chancellor was at elected*	Low contrast visual display	—	**LAN**	**P600**
				High contrast visual display	**LAN**	—	**P600**
[8]	**Gunter & Friederici** *[c,m,n]*	1999	Die weißen Zähne wurden vom Kind ✓**geputzt** *The white teeth were [by the] child brushed*				
			Die weißen Zähne wurden vom ***geputzt** *The white teeth were [by the] brushed*		—	**N400**	**P600**
			Die weißen Zähne wurden vom Kind ***putzen** *The white teeth were [by the] child brush*		—	**N400**	**P600**
[9]	**Federmeier et al.** *[c]*	2000	John wanted to ✓**eat**…				
			John wanted the * **eat**…		—	**N400**	**P600**
			John wanted the ✓**beer**…				
			John wanted to * **beer**…		—	**N400**	**P600**
[10]	**Hinojosa et al.** *[c,g,i,j]*	2003	La prueba ocultada por el fiscal ✓**apareció** *The proof hidden by the public prosecutor (it) appeared*				
			La prueba ocultada por el * **apareció** *The proof hidden by the appeared*	PS	**???**	**LAN/N400**	**P600**
			La prueba ocultada por el fiscal * **aparecí** *The proof hidden by the public prosecutor (I) appeared*	MS	—	**LAN**	**P600**

[11]	**Hagoort et al.** (Exp 1 & 2) *[c,f,o]*	2003	De houthakker ontweek de ijdele ✓**schroef** op dinsdag *The lumberjack dodged the vain propeller on Tuesday* De houthakker ontweek de ijdele * **schroeft** op dinsdag *The lumberjack dodged the vain propelled on Tuesday*		**—**	**AN**	**P600**
[12]	**Friederici & Meyer** *[c]*	2004	Er meinte dass Lisa ✓**Ärger** verursacht *He mentioned that Lisa trouble causes* Er meinte auch Lisa * **Ärger** verursacht *He mentioned also Lisa trouble causes*		**—**	**AN** 380-450	**P600**
[13]	**Frisch, Hahne, Friederici** (Exp 1) *[c,m]*	2004	Im Garten wurde oft ✓**gearbeitet** und… *In the garden was often worked and…*				
			Im Garten wurde am * **gearbeitet** und… *In the garden was on-the worked and…*	PS	**—**	**Positivity**	**P600**
			Der Garten wurde oft * **gearbeitet** und… *The garden was often worked and…*	AS	**—**	**N400**	**P600**
			Der garten wurde am * **gearbeitet** und… *The garden was on-the worked and…*	PS + AS	**—**	**—**	**P600**
[14]	**Martin-Loeches et al.** [1] *[c,o,r]*	2005	El compositor ✓**editó** la ópera *The composer edited-*[3rd per] *the opera.*				
			El compositor * **edición** la opera *The composer edition the opera*	PS	**—**	**LAN**	**P600**
			El compositor * **edité** la opera *The composer edited-*[1st per] *the opera*	MS	**—**	**LAN**	**P600**
[15]	**Lau et al.** [2] *[a,b,q]*	2006	*Although Erica kissed Mary's mother …* …she did not kiss Dana's * **of** the bride	+ellipsis			
			Although the bridesmaid kissed Mary … …she did not kiss Dana's * **of** the bride	-ellipsis	**LAN** 200-400		**—**
[16]	**Newman et al.** *[c,g,i]*	2007	The scientist criticized Max's proof ✓**of** the theorem The scientist criticized Max's * **of** proof the theorem		**?**	**LN**	**P600**
[17]	**Yamada & Neville** *[a-c,i,k,m,o]*	2007	Mommy can cut the meat with ✓**that** knife Mommy can cut the meat with her * **that** knife		**N100 /AN**	**—**	**P600**
[18]	**Roehm & Haider** [3] Exp 1 *[c,i]*	2009	Den Antrag bearbeiten ✓**sie** dennoch nicht *The application work-on-*[inf or 3rd/pl/pres] *they-*[3rd/pl] *anyhow not* Den Katalog abbestellen ***sie** aber bestimmt *The catalogue de-order-*[inf] *they-*[3rd/pl] *but surely*		**AN** 270-370	**—**	**P600**
	Exp 2		Den Antrag bearbeiten ✓**alle** selbst *The application work-on-*[inf or 3rd/pl/pres] *all-*[3rd/pl] *self* Den Katalog abbestellen ***alle** selbst *The catalogue de-order-*[inf] *all-*[3rd/pl] *self*		**—**	**PN** 340-440	**P600**
[19]	**Steinhauer et al.** *[f,i,r]*	2010	The scientist criticized Max's proof ✓**of** the theorem The scientist criticized Max's * **of** proof the theorem		**—**	**LN**[4]	**P600**

[20]	**Zhang et al.** [5] (Chinese) Exp 1	2010	Wei Li/ *ba*/ fresh/ pears/ slowly/ ✓**peeled**/ *le*/ two *Wei Li peeled the two fresh pears slowly*				
			Wei Li/ *ba*/ fresh/ pears/ slowly/ * **intimidated**/ *le*/ two *Wei Li intimidated two fresh pears slowly*	SEM	**—**	**N400**	**—**
			Wei Li/ *ba*/ fresh/ pears/ slowly/ * **knife**/ *le*/ two *Wei Li knife two fresh pears slowly*	PS	**—**	**LAN**	**—**
			Wei Li/ *ba*/ fresh/ pears/ slowly/ ** **piano**/ *le*/ two *Wei Li piano two fresh pears slowly*	PS+SEM	**—**	**LAN/ N400**	**P600**
	Exp 2 *[m]*		The girl/ bought/ *le*/ ✓**skirt**/ and/ glove *The girl bought a skirt and gloves*				
			The girl/ ate/ *le*/ * **skirt**/ and/ gloves *The girl ate a skirt and gloves*	SEM	**—**	**N400**	**—**
			The girl/ bought/ *le*/ *hen*/ * **skirt**/ and/ glove *The girl bought a very skirt and gloves*	PS	**N** 100-200	**LN**	**P600+AN**
			The girl/ ate/ *le*/ *hen*/ ** **skirt**/ and/ gloves *The girl ate a very skirt and gloves*	PS+SEM	**N** 100-200	**LN/N400**	**P600+AN**
[21]	**Steinhauer, White, Genesee**	inPrep	The man hoped to ✓**enjoy** the meal with friends The man made the ✓**meal** to enjoy with friends The man hoped to * **meal** the enjoy with friends The man made the * **enjoy** to meal with friends		**—**	**N400/LAN**	**P600**

[1] Martin-Loeches et al. (2005) also included, in addition to these three "short" conditions, six other "long" ones involving subject (3 conditions) vs. object (3 conditions) relative clauses. While both word-class and morpho-syntactic violations elicited P600 effects in these conditions, none elicited a LAN (and one condition – the word class violation in the subject relative clause condition – elicited a clear N400 effect prior to the P600). We refer the reader to their paper for discussion of these patterns. The important finding, for our purposes here, is that where the LAN effect was present, it was indistinguishable (either in terms of latency or topography) across the word category and morpho-syntactic violation types.

[2] Lau et al. (2006) did not include any grammatical control condition for this comparison. Strictly speaking this is a context manipulation design, though note that the manipulation kept all of the preceding main clause material constant. We refer readers to their paper for discussion. However it is worth mentioning here that their reported effect resembled neither the N125 nor the later 300–500 ms left/temporal effect typically found for the Neville et al. (1991) stimuli. Thus it is unclear how Lau et al.'s finding relates to previous findings using this type of local word order violation.

[3] Roehm & Haider's manipulation is labeled differently here than in their paper, where they marked the 1st verb these conditions as finite. Here we have marked them as either ambiguous between a finite or infinitival form, as in the correct conditions for both experiments, or unambiguously infinitival (as in the violation conditions of both experiments). The logic of their labeling seems to have been to consider how the parser would regard the relevant verbs *after* downstream disambiguating information was encountered (e.g., in the case of the violations, the downstream information indicates that the 1st verb should have been finite, but this is inconsistent with the intrinsic properties of these particular (particle verbs), which can only have the form depicted if they are infinitival). We refer readers to their paper for discussion of the broader interest of their manipulations (including other conditions they tested which are not included here).

[4] Steinhauer et al. (2010) label this left lateralized negativity as an "LTN" for "left temporal negativity", based on the fact that (as with Neville et al., 1991 and Newman et al., 2007) the effect had a left temporal maximum. Here we more neutrally label it as a left negativity (LN).

[5] Zhang et al. (2010), in addition to offering evidence of that syntactic violations do not block semantic N400 effects, also illustrate another important point. Note their Exp. 1 used a target manipulation design, and Exp. 2 used a context manipulation design. And, while the expected syntactic effects (ELAN) did not materialize in Exp. 1 (as with *all* other target manipulation designs), they did find an early negativity (though broadly distributed) for the syntactic and double/combined violations in Exp. 2. That baseline concerns played a role here is further suggested by the different way baselines were handled by these authors across the two studies (−100 to 0 in the target manipulation design in Exp. 1, and a post-stimulus onset 0–100 ms baseline in the context manipulation design in Exp. 2).

Table A2
Auditory studies.

#	Study [main text]	Year	Example stimuli	CONDITIONS	Phase I (100-300 ms)	Phase II (300-500)	Phase III (> 600)
[1]	**Friederici et al.** *[b,f,g,j,k,p]*	1993	De Finder wurde ✓**belohnt** *The finder was rewarded*				
			Das Parkett wurde * **bohnere** *The parquet was polish*	MS	**—**	**LAN**	**P600** (weak)
			Der Freund wurde im * **besucht** *The friend was in the visited*	PS	**LAN**	**LAN**	**—**
[2]	**Friederici, Hahne, Mecklinger** (Exp.1) *[e,j,l]*	1996	Das Metall wurde ✓**veredel-t** von dem Goldschmied den man… *The metal was refined by the goldsmith who was…*				
			Das Metall wurde zur ***veredel-t** von dem Goldschmied den man… *The metal was for refined by the goldsmith who was…*	PS (suffix)	**LAN** 400-600[1]		**Late LAN + P600**
[3]	**Hahne & Friederici** *[a,b,e,j-l,p]*	1999	Das Baby wurde ✓**gefüttert** *The baby was fed.*				
			Die Gans wurde im * **gefüttert** *The goose was in the fed*	*20% violations* *80% violations*	**LAN** **(L)AN**	**—** ??	**P600** **—**
[4]	**Hahne & Jescheniak** *[j,k,p]*	2001	Der Apfel wurde ✓**gepflückt** *The apple was being plucked* Die Birne wurde im * **geflückt** *The pear was being in-the plucked*	German PS	**AN**	**—**	**P600**
			Die Glabbe wurde ✓**gerottert** *The wibon was being rished* Das Fiehm wurde im * **gerottert** *The ploker was being in-the rished*	Jabberwocky PS	**AN**	**—**	**P600**
[5]	**Hahne & Friederici** *[e,j,k,m,p]*	2002	Das Brot wurde ✓**gegessen** *The bread was eaten*				
			Der Vulkan wurde * **gegessen** SEM *The volcano was eaten*	*Acceptability* *Sensicality*	— —	**N400** **N400**	— —
			Das Eis wurde im * **gegessen** PS *The ice cream was in-the eaten*	*Acceptability* *Sensicality*	**AN** **AN**	— —	**P600** **P600**
			Das Türschloß wurde im * **gegessen** PS+SEM *The door lock was in-the eaten*	*Acceptability* *Sensicality*	**AN** **AN**	— **N400**	**P600** —
[6]	**Frisch, Hahne, Friederici** (Exp 2) *[c,m]*	2004	Im Garten wurde oft ✓**gearbeitet** und… *In the garden was often worked and…*				
			Der Garten wurde oft * **gearbeitet** und ... *The garden was often worked and…*	AS	**—**	**N400**	**P600**
			Im Garten wurde am * **gearbeitet** und… *In the garden was on-the worked and…*	PS	**—**	**Positivity**	**P600**
			Der garten wurde am * **gearbeitet** und… *The garden was on-the worked and…*	PS + AS	**—**	**—**	**P600**
[7]	**Friederici, Gunter, Hahne, & Mauth** *[e,l-n,q]*	2004	Der Strauch wurde ✓**verpflanzt** von einem Gärtner, … *The bush was replanted by a gardener…*				
			Das Buch wurde * **verpflanzt** von einem Verleger,… *The book was replanted by a publisher,…*	SEM	**—**	**N400**	
			Der Strauch wurde trotz * **verpflanzt** von einem Gärtner,… The bush was despite replanted by a gardener,…	PS	**LAN** 350-600[1]		**P600**
			Das Buch wurde trotz ****verpflanzt** von einem Verleger,… *The book was despite replanted by a publisher,…*	PS+SEM	**LAN** 350-600[1]		**P600**
[8]	**Van den Brink & Hagoort** *[k,m-o,q]*	2004	Het vrouwtje veegde de vloer met een oude ✓**bezem** … *The woman swept the floor with an old broom …*				
			Het vrouwtje veegde de vloer met een oude ** **bedelde** … *The woman swept the floor with an old begged …*	PS+SEM	**—**	**N400 < AN**[2]	**P600**
			Het vrouwtje veegde de vloer met een oude ****klieder-de** … *The woman swept the floor with an old messed …*	PS+SEM	**—**	**N400 < AN**[2]	**P600**

Table A2
Auditory studies.

[9]	**Rossi et al** [3] *[d-h,j,k,m,p,q,s]*	2005/ (2006)	Der Junge im <u>Kindergarten</u> ✓**singt** ein Lied					
			The boy in-the kindergarten sings a song					
			Der Junge im <u>Kindergarten</u> * **singst** ein Lied	MS		**—**	**LAN**	**P600**
			The boy in-the kindergarten sing a song					
			Der Junge <u>im</u> * **singt** ein Lied	PS		**LAN**	**—**	**P600**
			The boy in-the sings a song					
			Der Junge <u>im</u> ** **singst** ein Lied	PS+MS		**LAN**	**—**	**P600**
			The boy in-the sing a song					
[10]	**Eckstein & Friederici** *[d,f,k]*	2006	Maria weiß, dass der Renter im <u>Alter</u> ✓**kränkelt**					
			Maria knows that the pensioner in-the seniority ails					
			Maria weiß, dass der Renter im <u>altert</u> * **kränkelt**			**LN** 200-400		**P600**
			Maria knows that the pensioner in-the grows-old ails					
[11]	**Ye et al.** *[h,j,k,m,p,q]*	2006	Shejishi zhizuo xinji, ba <u>buliao</u> ✓**cai**-le					
			The stylist to make new dresses, BA the cloth cut					
			Famugong kaicai senlin, ba <u>songshu</u> * **cai**-le	SEM		**N400** 150-400		**—**
			The timberjack exploiting the forest, BA pine trees cut					
			Shejishi zhizuo xinji, <u>ba</u> * **cai**-le	PS		**AN** 50-1000		**—**
			The stylist to make new dresses, BA cut			N400? 300-500		
			Famugong kaicai senlin <u>ba</u> ** **cai**-le	PS+SEM		**AN** 50-1000		—
			Exploiting the forest, the timberjack cut pine trees			N400? 300-500		
[12]	**Hestvik et al.** [4]	2007	The zebra said that the hippo <u>kissed</u> ✓**the camel** on the nose…					
			The zebra that the hippos <u>kissed</u> * **the camel** on the nose…			**LAN**		**P600**
			The zebra that the hippo <u>kissed</u> ✓**on** the nose…					
[13]	**Isel et al.** [5] *[h]*	2007	L'enfant qui est dans la <u>maison</u> ✓**dort**					
			The child who is in the house is sleeping					
			Le caillou qui est dans la <u>piscine</u> * **dort**	SEM			**N400**	
			The stone which is in the swimming-pool is sleeping					
			Le chauffeur qui est dans <u>la</u> * **dort**	PS		**LAN** (150-600)		**P600**
			The driver who is in the is sleeping					
			Le fauteuil qui est dans <u>la</u> ** **dort**	PS+SEM		**LAN** (150-600)		**P600**
			The chair which is in the is sleeping					
[14]	**Hastings & Kotz** *[a,d,k,o-q]*	2008	<u>er</u> ✓**kegelt** / <u>du</u> ✓**kegelst**					
			he bowls / you bowl					
			<u>er</u> * **kegelst** / <u>du</u> ***kegelt**	AGR	*Judgment task*	**(A)N**		**P600**
			he bowl / you bowls		*Visual distraction*	**(A)N**		**—**
			<u>er</u> ✓**kegelt** / <u>ein</u> ✓**Kegel**					
			he bowls / a cone					
			<u>er</u> * **Kegel** / <u>ein</u> * **kegelt**	PS	*Judgment task*	**(A)N**	**(A)N**	—
			he cone / a bowls		*Visual distraction*	**(A)N**	**(A)N**	—
[15]	**Pakulak & Neville** *[j,k,p,s]*	2010	Mommy can cut the meat <u>with</u> ✓**that** knife					
					High proficiency	**LAN**		**P600**
			Mommy can cut the meat with <u>her</u> * **that** knife		Low proficiency	**AN**		**P600**

[1] These effects are measured post target word *onset*. See our main text discussion of issues related to how stimuli are time-locked to EEG signals.
[2] The onset of the N400 effect in this study preceded that of the AN (i.e., the "<" refers to the N400 having a shorter onset latency).
[3] Results reported here are for native speakers only (Rossi et al., 2005), though note that Rossi et al. (2006), where non-native speakers were tested, is also discussed in the main text.
[4] Hestvik et al., (not discussed in the main text) can be considered a context manipulation approach which aimed to include an additional control. Their first two conditions differ in whether the target NP is contained in a complement clause (correct), or a relative clause (where a gap is expected following the underlined verb <u>*kissed*</u>). Though these conditions are matched in terms of the (here: 5) preceding lexical elements, this contrast is confounded with expected effects tied to (object) relative clause processing, namely working memory related (left/anterior) negativities. To control for this possibility, they also included the third condition which should share the working memory negativity with the violation condition, but which should not yield any violation effect following the underlined verb (<u>*kissed*</u>). Thus this design could be referred to as a "controlled context manipulation" paradigm. However, the findings of this study, in our view, are not straightforwardly interpretable (the third "control" condition also differs from the other two in having a prepositional phrase following the verb, as opposed to an NP in the other two cases). Another difficulty with these contrasts relates to the issue concerning "outright violations". Here, the determiner following the relevant verb (<u>*kissed*</u>), like many other studies (see Section 2.1.2), could be legally continued (as the authors note, e.g., *The zebra that the hippo kissed the camel for*, or *The zebra that the hippo kissed the nose of*, etc., see their p312). While these latter continuations can be viewed as rather rare constructions, another grammatical continuation cannot: *The zebra that the hippo kissed the other day*… We refer readers to their paper for further discussion.
[5] Experiment 1. Note Experiment 2 included only syntactic violations and replicated the Experiment 1 pattern.

Index	Section		
[a]	**1.**		**Introduction**
	2.		**Critique**
[b]	**2.1**		**Problem #1: empirical concerns**
[c]		2.1.1	Modality independence of ELANs
[d]		2.1.2	ELAN specificity for "outright" violations
[e]		2.1.3	Strategies, seriality, and temporal availability of information
	2.2		**Problem #2: Methodological Issues (beyond the face value)**
[f]		2.2.1	Two types of PS violation paradigms
[g]		2.2.2	A note on (a)symmetry in violation paradigm designs
[h]		2.2.3	Context effects in ERPs: spillover and offset
[i]		2.2.4	Context/baseline problems in ELAN reading studies
[j]		2.2.5	Context/baseline problems in auditory ELAN studies
[k]		2.2.6	Local ELANs, sustained negativities, and P600s
[l]		2.2.7	Local ELAN-like negativities without P600s
[m]	**2.3**		**Semantic blocking as evidence for "syntax-first" models**
[n]		2.3.1	Friederici et al. (2004)/van den Brink and Hagoort (2004)
		2.3.2	Explanations
		2.3.3	An alternative
	3		**New landscape**
[o]	**3.1**		**PS violations in reading**
[p]	**3.2**		**PS violations in speech**
[q]		3.2.1	Sustained negativities
[r]		3.2.2	Working memory and sustained negativities
[s]	**3.3**		**ELANs in language acquisition and breakdown**
	4.		**Conclusion: What does all this leave us with?**

References

Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language, 38*(1), 112–130.

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

Bornkessel-Schlesewky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research, 59*, 55–73.

Bornkessel, I., & Schlesewsky, M. (2009). *Processing syntax and morphology. A neurocognitive perspective*. Oxford University Press.

Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open-and closed-class words. *Journal of Cognitive Neuroscience, 11*(3), 261–281.

Capek, C. M., Grossi, G., Newman, A. J., McBurney, S. L., Corina, D., Roeder, B., et al. (2009). Brain systems mediating semantic and syntactic processing in deaf native signers: Biological invariance and modality specificity. *PNAS, 106*(21), 8784–8789.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of processing. *Psychophysiology, 32*(3), 274–285.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564–570.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Dikker, S., Rabagliati, H., & Pylkkanen, L. (2009). Sensitivity to syntax in the visual cortex. *Cognition, 110*, 293–321.

Donchin, E. (1981). Surprise! Surprise? *Psychophysiology, 18*, 493–513.

Drury, J. E., Steinhauer, K., Pancheva, R., & Ullman, M. T. (in preparation). (In)definiteness ERP effects in English existential constructions.

Drury, J. E., & Steinhauer, K. (2009). Brain potentials for logical semantics/pragmatics. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: from experiment to theory*. Palgrave Macmillan.

Drury, J. E., Ullman, M. T., & Steinhauer, K. (2010). Show us the baseline: On syntactic ELAN effects in ERP reading studies. *Journal of Cognitive Neuroscience Supplement*, 267.

Dwivedi, V., Phillips, N. A., Lague-Beauvais, M., & Baum, S. R. (2006). An electrophysiological study of mood, modal context, and anaphora. *Brain Research, 1117*(1), 135–153.

Eckstein, K., & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research, 25*, 130–143.

Eckstein, K., & Friederici, A. (2006). It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience, 18*(10), 1696–1711.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123*, 2552–2566.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research, 30*(3), 321–338.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250–272.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: The MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge: MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance*. Hillsdale, NJ: Lawrence Erlbaum.

Friederici, A. D. (1990). On the properties of cognitive modules. *Psychological Research, 52*(2-3), 175–180.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language, 50*, 259–281.

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

Friederici, A. D., & Frisch, S. (2000). Verb-argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*, 476–507.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). The temporal structure of syntactic parsing: Early and late effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1219–1248.

Friederici, A. D., Hahne, A., & von Cramon, D. Y. (1998). First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: Electrophysiological evidence for a double dissociation. *Brain and Language, 62*(3), 311–341.

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.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory and Cognition, 27*(3), 438–453.

Friederici, A., & Kotz, S. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20*, 8–17.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*(2), 305–323.

Friederici, A., & Meyer, M. (2004). The brain knows the difference: Two types of grammatical violations. *Brain Research, 1000*, 72–77.

Friederici, A., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantics interface. *Brain Research, 1146*, 50–58.

Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *NeuroReport, 15*(1), 165–169.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, 99*(1), 529–534.

Frisch, S., Hahne, A., & Friederici, A. (2004). Word category and verb–argument structure information in the dynamics of parsing. *Cognition, 91*, 191–219.

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology, 36*(1), 126–137.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34*(6), 660–676.

Gunter, T. C., Friederici, A., & Hahne, A. (1999). Brain responses during sentence reading: visual input affects central processes. *Neuroreport, 10*(15), 3175–3178.

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage, 20*, S18–S29. doi:10.1016/j.neuroimage.2003.09.013.

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

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439–483.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience, 16*(7), 1302–1318.
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.
Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research, 11*, 199–212.
Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*(3), 339–356.
Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*(3), 251–266.
Hasting, A. S., & Kotz, S. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience, 20*(7), 1207–1219.
Hinojosa, J. A., Martin-Loeches, M., Casado, P., Munoz, F., & Rubia, F. J. (2003). Similarities and differences between phrase structure and morphosyntactic violations in Spanish: An event-related potentials study. *Language and Cognitive Processes, 18*(2), 113–142.
Hwang, H., & Steinhauer, K. (2010). Prosodic phrasing in spoken Korean garden path sentences: An ERP study. *Journal of Cognitive Neuroscience* (Suppl. 2010).
Hwang, H., & Steinhauer, K. (in press). Phrase length matters: The interplay between implicit prosody and syntax in Korean 'garden path' sentences. *Journal of Cognitive Neuroscience*. <http://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_00001>.
Hwang, H., & Steinhauer, K. (in preparation). Prosodic effects of phrase length in Korean spoken garden-path sentences: An ERP study.
Isel, F., Hahne, A., Maess, B., & Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology, 74*, 337–346.
Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., & Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport, 21*, 8–13.
King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*(3), 376–395.
Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERP's on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196–214.
Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.
Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*(12), 463–470.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(1), 203–205.
Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)Constructing the N400. *Nature Review Neuroscience, 9*, 920–933.
Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*, 74–88.
Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1–71.
Martin-Loeches, M., Munoz, F., Casado, P., Melcon, A., & Frenandez-Frias, C. (2005). Are the anterior negativities to grammatical violations indexing working memory? *Psychophysiology, 42*, 508–519.
McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes, 4*(SI), 287–335.
Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. B. R. D. Luce & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley.
Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience, 17*, 1229–1244.
Münte, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217–226.
Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395*(6697), 71–73.
Neville, H., 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.
Neville, H. J., Mills, D. J., & Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex2*, 244–258.
Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *Neuroimage, 34*, 435–445.
Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: event related potentials as real time reflections of sentence comprehension and language learning. In Carreiras, M. & Clifton, J. (Eds.), *The on-line study of sentence comprehension: eye tracking*. ERP and Beyond.
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.
Pakulak, E. (2008). An investigation of the effects of proficiency and age of acquisition on neural organization for syntactic processing using ERPs and fMRI. Unpublished Doctoral dissertation, University of Oregon, Eugene.
Pakulak, E., & Neville, H. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22*(12), 2728–2744.
Patel, Aniruddh D., Gibson, Edward, Ratner, Jennifer, Besson, Mireille, & Holcomb, Phillip J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717–733.
Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience, 23*(10), 2731–2751.
Pauker, E., & Steinhauer, K. (2010). The early left anterior negativity (ELAN): A reliable marker of phrase structure violations, or a context-driven artifact? *Annual CUNY conference on Human Sentence Processing*, NYA (NY), USA, March 2010.
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233–238.
Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*(3), 407.
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*, 245–255.
Roehm, D., & Haider, H. (2009). Small is beautiful: The processing of the left periphery in German. *Lingua, 119*, 1501–1522.
Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030–2048.
Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters, 384*, 228–233.
Ruchkin, D. S., Johnson, R., Jr., Grafman, J., Canoune, H. L., & Ritter, W. (1992). Distinctions and similarities among working memory processes: An event-related potential study. *Cognitive Brain Research, 1*, 53–66.
Sabisch, B., Hahne, A., Glass, E., von Suchodoletz, W., & Friederici, A. D. (2006). Auditory language comprehension in children with developmental dyslexia: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 18*(10), 1676–1695.
Sakai, K. L., Hashimoto, R., & Homae, F. (2001). Sentence processing in the cerebral cortex. *Neuroscience Research, 39*(1), 1–10.
Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer & H. Whitaker (Eds.), *Handbook of the cognitive neuroscience of language* (pp. 91–104). New York: Elsevier.
Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., & Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia, 48*, 1525–1542.
Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*, 13–41.
Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191–196.
Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2*, 717–726.
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*, 158–171.
Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*, 967–985.
van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*(6), 1068.
Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning Memory & Cognition, 25*, 394–417.
Vos, Sandra H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001a). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes, 16*(1), 65–103.
Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001b). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38*(1), 41–63.
Yamada, Y., & Neville, H. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research, 1130*, 167–180.
Ye, Z., Luo, Y., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research, 1071*, 186–196.
Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 765–781.