Language and Cognitive Processes

Electrophysiological correlates of morphosyntactic integration in German phrasal context

D. J. Davidson [a] , A. Hanulíková [a b] & P. Indefrey [c d]

[a] Basque Centre on Cognition, Brain and Language, Donostia,
Spain

[b] Max Planck Institute for Psycholinguistics, Nijmegen, The
Netherlands

[c] Abt. für Allgemeine Sprachwissenschaft, Institut für Sprache
und Information, Heinrich-Heine-Universität, Düsseldorf,
Germany

[d] Donders Centre for Cognitive Neuroimaging, Donders Institute
for Brain, Cognition, and Behavior, Nijmegen, The Netherlands

Available online: 03 Feb 2012

PLEASE SCROLL DOWN FOR ARTICLE

arising directly or indirectly in connection with or arising out of the use of this
material.

CNL Electrophysiological correlates of morphosyntactic integration in German phrasal context

D. J. Davidson[1], A. Hanulíková[1,2], and P. Indefrey[3,4]

[1]Basque Centre on Cognition, Brain and Language, Donostia, Spain
[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[3]Abt. für Allgemeine Sprachwissenschaft, Institut für Sprache und Information, Heinrich-Heine-Universität, Düsseldorf, Germany
[4]Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition, and Behavior, Nijmegen, The Netherlands

The morphosyntactic paradigm of an inflected word can influence isolated word recognition, but its role in multiple-word phrasal integration is less clear. We examined the electrophysiological response to adjectives in short German prepositional phrases to evaluate whether strong and weak forms of the adjective show a differential response, and whether paradigm variables are related to this response. Twenty native German speakers classified serially presented phrases as grammatically correct or not while the electroencephalogram (EEG) was recorded. A functional mixed effects model of the response to grammatically correct trials revealed a differential response to strong and weak forms of the adjectives. This response difference depended on whether the preceding preposition imposed accusative or dative case. The lexically conditioned information content of the adjectives modulated a later interval of the response. The results indicate that grammatical context modulates the response to morphosyntactic information content, and lends support to the role of paradigm structure in integrative phrasal processing.

Keywords: Paradigm; Syntactic integration; Parsing; Information theory; Inflectional morphology.

INTRODUCTION

Recent approaches to the recognition of inflected words have employed a statistical framework based on information theory and word-and-paradigm morphology (Milin, Kuperman, Kostić, & Baayen, 2008). In this approach, the speed of a classification

response to a word in a lexical decision task is inversely related to the relative amount of information provided by that word, estimated from a probability model. This probability is taken from a frequency estimate of the word relative to a corpus sample of a given size, or by more complex measures of the probability distribution associated with a word. The experiments based on this approach have shown that paradigm-specific information content influences the response to words that have been presented in isolation. However, much less is known about how paradigm-related information content influences morphosyntactic integration in phrasal (syntagmatic) context. This issue is relevant for paradigm-based models because the choice of morphological suffix in an inflected word is a reflection of the grammatical context in which it occurs (see section below entitled "Information Content of Morphosyntactic Paradigms"). If the paradigm-specific information content is used during phrasal processing, this would add further support for the proposal that paradigmatic organisation is relevant for comprehension. More generally, there has been increased interest in the application of information theory to sentence processing (Levy, 2008), but the specific interaction of the paradigm distribution and grammatical integration has yet to be addressed.

Here, we use electrophysiology to examine how the distribution of an inflected form within a paradigm affects the response to the strong and weak adjectival declension in German. In the remainder of the introduction, we first review the existing literature on variables related to lexical paradigms. Then we describe how electroencephalography (EEG) has been used to examine morphosyntactic recognition in phrasal contexts, and finally we summarise and propose several hypotheses.

INFORMATION CONTENT OF MORPHOSYNTACTIC PARADIGMS

Several studies have now shown that morphological paradigms can influence isolated word recognition in visual lexical decision tasks. Kostić (1991, 1995) showed that a statistical measure that incorporated the frequency of an inflected form within a paradigm, as well as the number of grammatical functions or meanings of a word, was highly positively correlated with lexical decision times during the recognition of Serbian nouns. This measure was based on information theory, indicating the relative amount of information that an inflection suffix provides, relative to its paradigm. Moscoso del Prado Martín, Kostić, and Baayen (2004) found that lexical decision times for Dutch nouns were positively correlated with inflectional entropy. Inflectional entropy increases in a paradigm when there are more inflectional variants possible, and/or when the variants have similar probabilities. Baayen, McQueen, Dijkstra, and Schreuder (2003) showed that the relative frequency of inflected forms of Dutch nouns influenced visual and auditory lexical decision times. Milin et al. (2008) reviewed more recent evidence supporting the role of information content and entropy in lexical decision. Also note that the existing literature suggests important differences between grammatical categories. Kostić and Katz (1987) compared lexical decision times for isolated Serbian nouns, adjectives, and verbs and found that, unlike nouns, responses to adjectives were not fastest to most common default nominative case forms, and secondly, that surface frequency predicted response times. The results of these studies indicate that paradigm distribution can influence isolated word recognition, but it is not clear that the effects are the same across grammatical categories.

Less is known about whether there is an influence of morphological paradigms in phrasal constructions, although there is good evidence that phrasal context has a significant impact on morphosyntactic processing. For example, an extensive series of

experiments with Serbian has shown that inflected words are recognised more quickly (as revealed by lexical decision times) in correct grammatical contexts as compared to grammatical-violation contexts (Gurjanov, Lukatela, Lukatela, Savić, & Turvey, 1985; Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985; Lukatela, Kostić, Feldman, & Turvey, 1983; Lukatela, Kostić, Todorović, Carello, & Turvey, 1987). Also, related to the present study, Bölte and Connine (2004) have shown that correct inflection facilitates the recognition of German nouns when the noun is preceded by a determiner, also using a lexical decision task. However, recently Hyönä, Vainio, and Laine (2002) have shown that lexical factors which operate in isolated lexical decision tasks do not necessarily have the same effect in sentence context (in Finnish), and Vainio, Hyönä, and Pajunen (2008) have shown relatively delayed effects of grammatical agreement during sentence reading, in that the facilitatory effects of agreement emerged in total reading time measures rather than first-pass fixation measures. Thus, it is an open question whether the lexical factors identified in earlier work using isolated-word lexical decision also operate during phrasal processing, and if so, when.

There is some evidence that paradigm organisation can determine word-level processing in German, but the question remains whether it affects phrase-level processing. Clahsen, Eisenbeiss, Hadler, and Sonnenstuhl (2001) have shown that paradigm specificity influences repetition priming patterns in German adjectives and verbs. However, it is not yet clear how repetition priming maps onto phrasal composition, or whether similar specificity results would be obtained in phrasal contexts (also see Barber & Carreiras, 2005). Penke, Janssen, and Eisenbeiss (2004) found evidence consistent with an under-specified morphological paradigm for German adjectival paradigms in a sentence matching task, but they did not investigate the influence of the item-specific probability distribution for the paradigm. Unlike the behavioral results for lexical decision, the tasks that have been used for sentence contexts so far do not allow for an estimate of when (within the sentence) the effect of paradigm distribution would be expected to operate.

PROBABILITY MODEL FOR THE GERMAN STRONG AND WEAK ADJECTIVAL PARADIGMS

In the experiment reported below, the influence of the paradigm probability distribution is investigated in German adjectives using a probability model derived from Milin et al. (2008). German determiners and adjectives show agreement with their head nouns for the features of case, gender, and number. The degree to which these features are expressed by an adjective suffix differs between three classes termed the strong, weak, and mixed declension. An adjective may take suffixes of all three classes, depending on the preceding elements of the noun phrase (NP). This dependency is considered to be a syntactic dependency rather than a semantic or phonological dependency (Zwicky, 1986) and, following Schlenker (1999), it can be described as a rule according to which syntactic features are to be expressed only on the first inflectable element of an NP. If the adjective is the first inflectable element, it takes on a strong suffix. The suffix –em in [1], for example, specifies dative case and nonfeminine gender. Please note that the examples are indexed with square brackets and equations are indexed with parentheses. By contrast, if the adjective is preceded by a definite determiner that expresses the feature information as in [2], the adjective has a weak suffix –en that is compatible with the feature specification of the determiner but does not express the features itself.

[1]   mit *kleinem* Boot   STRONG
        "with a small boat"

[2]   mit dem *kleinen* Boot   WEAK
        "with the small boat"

As a consequence of this, a greater variety of suffixes (*–e, –en, –em, –es, –er*) are used with the strong declension than the weak (*–e, –en*), as shown in Table 1.

The five German adjective suffixes show a high degree of syncretism expressing 72 possible combinations of case, gender, number, and declension class. Linguistic analyses have dealt with the syncretism by using a hierarchical feature specification, such that some suffixes are specified for case, number, and gender features and others are treated as default or "elsewhere" forms with reduced or no specification (Bierwisch, 1967; Blevins, 1995; Cahill & Gazdar, 1997; Clahsen, Eisenbeiss, Hadler, & Sonnenstuhl, 2001; Penke et al., 2004; Schlenker, 1999; Wunderlich, 1997; Zwicky, 1986). Commonly assumed features in analyses of German determiner and adjective morphology are $\pm$OBLIQUE and $\pm$GOVERNED for case (nominative$_{-OBL\ -GOV}$, accusative$_{-OBL\ +GOV}$, dative$_{+OBL\ +GOV}$, and genitive$_{+OBL\ -GOV}$), $\pm$MASCULINE and $\pm$FEMININE for gender (masculine$_{+M\ -F}$, neuter$_{-M\ -F}$, and feminine$_{-M\ +F}$), and $\pm$PLURAL for number (singular$_{-PL}$, plural$_{+PL}$). All analyses assume to some degree homophonic suffixes and in most cases the suffixes are analysed separately for strong and weak declension with the five suffixes of the strong declension carrying more feature specifications than the two suffixes of the weak declension. Following Schlenker's (1999) general idea of a left-to-right accumulation of features in the processing of a German noun phrase, the adjective suffix *–em* in example [1] specifies the features $+$OBL, $+$GOV, –F, and –PL. In example [2] the same features are specified by the determiner *dem* and the adjective suffix *–en* is a default suffix adding no feature information to the NP. Note that due to the syncretism of the strong suffix *–em*, gender remains underspecified as nonfeminine in the dative case. In accusative case NPs such as [3], which is governed by the preposition *ohne*, the strong suffix *–es* underspecifies case but fully specifies neuter gender with the features –OBL, –M, –F, and –PL. In the corresponding definite NP [4] the determiner again specifies the same features. Even though in this case the weak adjective suffix *–e* specifies number as singular, from a processing view this is redundant and adds no further feature information to the NP. Note that this is different for NPs with the definite determiner *die* indicating feminine singular or plural of all genders. To keep the weak conditions homogeneous, we therefore only used masculine and neuter nouns in the present experiment.

[3]   ohne kleines Boot   STRONG
        "without a small boat"

[4]   ohne das kleine Boot   WEAK
        "without the small boat"

The pattern of weak and strong endings can be quantified by the respective entropy values of the strong and the weak paradigms. To illustrate this, Table 2 shows the

TABLE 1
Strong and weak adjective declension in German according to case (nominative, accusative, dative, genitive), gender (masculine, feminine, neuter), and number (singular, plural)

	Masc-sg	*Neu-sg*	*Fem-sg*	*-pl*
Nom	r/e	s/e	e/e	e/n
Acc	*n/n*	*s/e*	e/e	e/n
Dat	*m/n*	*m/n*	r/n	n/n
Gen	n/n	n/n	r/n	r/n

Note: Suffixes used in the current experiment shown in italics.

relative frequency F, estimated proportion Pr_p, and estimated information I_{w_x} for the stem + suffix combinations of an example adjective stem $w = klein$-, as well as for both the strong and weak adjective suffixes as such. The frequencies and proportions were obtained from a tagged corpus sample of German text (the Digitales Wörterbuch der deutschen Sprache-Kerncorpus, 2008; DWDS core corpus). The DWDS corpus consists of a text corpus of over 100 million words spanning the time period 1900–2000. The corpus contains text from five different genres (approximate percentage in parenthesis, see Geyken, 2007): journalism (27%), literary text (26%), science literature (22%), and other nonfiction (20%). The estimates were obtained by using a regular expression search over sequences of grammatical category tags corresponding to "Prep Adj Noun" for the strong or "Prep Det Adj Noun" for the weak. Please see www.dwds.de for more information about the composition of the corpus. The amount of information conveyed by the suffix x

$$I_x = -\log_2 Pr_\pi(x) \qquad (1)$$

is calculated from the probability of the suffix $Pr_\pi(x)$, estimated from the frequency of the suffixes $F(x)$ relative to the frequency of the suffixes in the paradigms π:

$$Pr_\pi = \frac{F(x)}{\sum_x F(x)} \qquad (2)$$

The distribution in Table 2 shows that the amount of information is more evenly distributed over the strong paradigm (the probabilities are more similar to each other) compared to the weak. In addition, the two suffixes that are used only in the strong paradigm, *–em* ($F(x) = 80{,}398$) and *–es* ($F(x) = 9{,}272$), are less common than the

TABLE 2
Probability distribution for the strong and weak inflected forms of the adjective *klein*- ($F(w_x)$, $Pr_\pi(w_x)$, I_{w_x}) in the prepositional phrase *Prep (Det) Adj Noun*, as well as the distribution of suffixes for the same phrases over all adjectives ($F(x)$, $Pr_\pi(x)$, I_x)

Class	w_x	$F(w_x)$	$Pr_\pi(w_x)$	I_{w_x}	$F(x)$	$Pr_\pi(x)$	I_x
Strong	klein-*en*	2,216	0.6149	0.7016	184,273	0.3523	1.50495
	klein-*e*	821	0.2278	2.1341	100,691	0.1925	2.37686
	klein-*em*	265	0.0735	3.7655	80,398	0.1537	2.70156
	klein-*er*	262	0.0727	3.7820	148,360	0.2837	1.81769
	klein-*es*	40	0.0111	6.4935	9,272	0.0177	5.81777
Weak	klein-*en*	2,888	0.8273	0.2736	371,216	0.7864	0.34661
	klein-*e*	600	0.1719	2.5406	97,548	0.2067	2.27468
	klein-*em**	3	0.0009	10.1845	93	0.0002	12.30935
	klein-*er**	0	0.0000		3,066	0.0065	7.26636
	klein-*es**	0	0.0000		103	0.0002	12.16201

Note: The first three numerical columns refer to the properties of the specific item *klein*-, while the last three numerical columns refer to the properties of the suffixes over all adjectives. Estimated frequencies taken from the tagged DWDS-Kerncorpus (www.dwds.de) comprising data over the years 1900–2000.
*Observed in the corpus for either the *klein*- or other stems, but normatively incorrect.

others. The entropy of the strong declension, 2.0218, is therefore higher than the weak, 0.7949, calculated with (3) over the suffixes in Table 2.

$$H = - \sum_x \mathrm{Pr}_\pi(w_x) \log_2 \mathrm{Pr}_\pi(w_x) \tag{3}$$

The entropy indices capture the intuition expressed by the descriptive terms "strong" and "weak" because the weak form is thought to convey less information, because it has already been expressed on the definite determiner in the phrase. Here, we wish to extend the previous work by Kostić and colleagues using lexical decision tasks to phrasal context. If language comprehenders are sensitive to the amount of information conveyed by a term in phrasal context, then there should be a larger response time and error cost associated with recognising the strong form relative to the weak.

In addition to the global probability model, lexeme-specific measures of information can be derived. The amount of information conveyed by a suffix combined with a particular stem

$$I_{x|w} = - \log_2 \mathrm{Pr}_\pi(w_x) + \log_2 \mathrm{Pr}_\pi(w) \tag{4}$$

is calculated from the probability of the inflected form $Pr_\pi(w_x)$ and $Pr_\pi(w)$. The first probability can be estimated from the frequency of the inflected form w_x relative to the sum of the frequencies of all forms of the stem's paradigm,

$$\mathrm{Pr}_\pi(w_x) = \frac{F(w_x)}{\sum_x F(w_x)} \tag{5}$$

while the second, $Pr_\pi(w)$, is estimated from the frequency of the adjectival stem w in the corpus (this is also termed lemma frequency by some authors).

$$\mathrm{Pr}_\pi(w) = \frac{F(w)}{N} \tag{6}$$

A related quantity is the relative amount of information conveyed by a particular stem + suffix combination relative to the number of word forms N in the corpus

$$I_{w_x} = - \log_2 \mathrm{Pr}_\pi(w_x) \tag{7}$$

which is derived from $Pr_N(w_x)$ (also sometimes termed the form frequency).

$$\mathrm{Pr}_N(w_x) = \frac{F(w_x)}{N} \tag{8}$$

The difference between $I_{x|w}$ and I_{w_x} is that $I_{x|w}$ models the amount of information relative to the paradigm, while I_{w_x} models information relative to the entire corpus.

Finally, we can calculate the relative entropy (Kullback–Lieber divergence, D) for each adjective as in (9), as an estimate for how much the distribution of suffixes in the strong and weak paradigms of a particular stem deviates from the distribution of suffixes in the strong and weak declension paradigms in general. Here, P is the probability distribution of the paradigm for the stem, and Q is the distribution of the

declension class (here strong and weak). The value of D for *klein-* for the strong declension would be 0.3207, and for the weak, 0.0145, using the values in Table 2 (replacing values of 0 in the table with 1), indicating greater divergence of the suffixes of the adjective *klein-* from the strong general paradigm than the weak general paradigm, for which D is close to 0 (no divergence).

$$D(P\|Q) = \sum_x \mathrm{Pr}_\pi(w_x) \log_2 \frac{\mathrm{Pr}_N(w_x)}{\mathrm{Pr}_N(x)} \tag{9}$$

ELECTROPHYSIOLOGICAL ACTIVITY RELATED TO MORPHOSYNTAX

It is possible that electrophysiological brain responses recorded with EEG may also be related to the variation in the different probability measures in a paradigm. This could provide additional information about when these variables have an influence on contextualised recognition, because EEG reflects some of the brain activity involved in recognition as it unfolds over time. However, it is not clear from the previous behavioral work using either single word lexical decision or sentence processing, which cortical systems are responsible for processing morphosyntax, or what the expected timing of the evoked response should be. It can be assumed based on previous lexical decision tasks that the morphological information is available in under 1 s, as most average behavioral responses (response times, fixation durations) in these tasks are well within this interval. However, because these behavioral measures capture only a limited number of data points related to activity per trial, the full profile of activity is not known. Also, behavioral measurements are limited to motor responses, which reflect the timing of motor system activity, and not directly the system of interest. In contrast, the EEG yields a more densely sampled time series, which reflects (some fraction of) the activity that is directly linked to the system of interest.

Electrophysiological studies employing contrasts between grammatical violations and controls (Barber & Carreiras, 2005; Gunter, Friederici, & Schriefers, 2000; Münte & Heinze, 1994) have shown several event-related potential (ERP) components responsive to morphosyntactic agreement. These include the left anterior negative component (LAN), as well as a late positive (P600) component. If, for instance, the syntactic agreement of gender between a noun and its preceding determiner is violated (*den*$_{acc,masc,sing}$*Land*$_{neut,sing}$), the LAN can be observed as early as 0.3 s after the onset of the word that creates a mismatch with the preceding context, and usually shows a left frontal maximum (but see Hagoort, Wassenaar, & Brown, 2003; Münte, Matzke, & Johannes, 1997). Early negativities have been observed to morphosyntactic violations in German (e.g., Gunter et al., 2000; Koester, Gunter, Wagner, & Friederici, 2004), Spanish (Barber & Carreiras, 2005), English (Münte, Heinze, & Mangun, 1993), and Hebrew (Deutsch & Bentin, 2001). In addition to a LAN, these studies report the P600 starting about 0.5 s after the onset of the violation with a maximum at posterior electrode locations. These two ERP components have been reported for violations that occur in both isolated phrases and sentences. The P600 is sometimes described as a "late" component, in that it follows the early (N100 or P200) or mid-latency (N400 or LAN) components, but it is understood to be elicited by the same critical word. There are studies suggesting that the late positivity can be found for processing syntactically incorrect structures (Hagoort, Brown, & Groothusen, 1993) as well as infrequent structures (Osterhout & Holcomb, 1992). The anterior negativities are often observed

for incorrect structures (Friederici, Hahne, & Mecklinger, 1996; Münte et al. 1993, 1997; Osterhout & Mobley, 1995) but have been observed for noncanonical structures as well (e.g., Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998).

EEG findings concerning the violation effects have motivated a cognitive model of processing for both noun phrase and verbal agreement that involves several stages of processing (Barber & Carreiras, 2005; Faussart, Jakubowitz, & Costes, 1999; Molinaro, Vespignani, & Job, 2008). In this view, there is an early stage of lexical access during which a word is selected based on the perceptual input. This is followed by a stage of lexical recognition, during which morphosyntactic or semantic features are accessed. Finally, depending on the previous phrasal context, context integration and possibly re-analysis occurs (following a violation). In the previous literature, the ERP violation effects described above have been used to determine whether some features such as person, gender, or number are accessed before others, or whether the violations of different features are re-analysed with different time courses. For the most part, the relative insensitivity of ERP components to different types of agreement violations has been taken as evidence against feature hierarchies during lexical recognition or integration (Barber & Carreiras, 2005; Silva-Pereyra & Carreiras, 2007), an issue which the present study does not address. However, the influence of different types of features on relatively late components (e.g., P600 or later, see also Vainio et al., 2008) has been taken as evidence for the stages of processing in this model (Molinaro et al., 2008). Interestingly, the late influence of features is taken to be a by-product of the differential access to the stem and suffix during re-analysis. This was taken to be consistent with the hypothesis proposed by Barber and Carreiras (2005), elaborated by Molinaro et al. (2008), that the timing of P600 responses with different latencies might reflect different stages of a re-analysis process. However, to date the probability distribution of the combined stem and suffix during grammatical phrasal processing has not figured prominently in this research. The present study can be seen as providing more information about the influence of these statistical factors on the lexical recognition and integration phases of the parsing of grammatical phrases during agreement processing.

Another line of comprehension research has examined so-called attraction effects in number violation sentences, mainly in English. Similar to the well-known attraction effect seen during production (Bock & Miller, 1991), subjects show an attenuated reading time disruption for violations of subject–verb number agreement when there is an intervening noun that agrees with the verb (Pearlmutter, Garnsey, & Bock, 1999), and this is also reflected in a reduced P600 amplitude (Kaan, 2002). Theoretical explanations of this data pattern have emphasised either mechanisms of feature transmission during the construction of the subject noun phrase, or alternatively, errors in retrieving the correct subject representation when processing the verb (see Wagers, Lau, & Phillips, 2009 for a recent review). Although the processes that are hypothesised to explain the attraction effects seen in the comprehension literature are similar to those used to explain grammatical violation effects in the electrophysiological literature (see e.g., Nevins, Dillon, Malhotra, & Phillips, 2007), it is important to note that the attraction effect occurs over a dependency spanning a longer distance than the local agreement effects studied in the present paper. Nevertheless, some of the processes might be common to different types of agreement. For example, Wagers et al. (2009) suggest that a plurality effect observed in several reading studies might have basis in morphological complexity (see their Table 12). Statistical measures of morphosyntactic complexity, such as those described here, might be informative in this regard.

More recently, lexical variables such as morphological family size have been considered in neurophysiological studies. Using an English visual lexical decision task, Solomyak and Marantz (2010) have recently shown with magnetoencephalography (MEG) that activity in a time window around $0.17 \, s$ is correlated with the conditional probability of a word given its suffix. This approach used a region-of-interest analysis with a source reconstruction to examine the time course of the brain response to the presentation of words presented as text in English. This result provides support for the approach of relating continuous lexical variables to single-trial electrophysiological responses. Effects of morphological family size have also been observed with MEG (Pylkkänen, Feintuch, Hopkins, & Marantz, 2004), but as with the behaviorally oriented work reviewed above, not in phrasal context. The present study compares the responses to strong and weak forms in phrasal contexts, and will focus mainly on grammatically correct phrases.

Other work with EEG supports the role of the amount of (semantic) information in phrasal processing. The original report of the N400 semantic violation effect in sentence processing (Kutas & Hillyard, 1980), as well as the work that followed, shows that words with high information content (e.g., low probability), as determined by the preceding semantic or discourse context, elicit a larger amplitude potential. In another example, Gunter et al. (2000) found the LAN + P600 complex to gender violations only when the closure probability of the noun is high. Low closure probability nouns elicited the LAN, but the P600 was strongly reduced. Other studies have argued that the P600 violation effect, seen in response to grammatically incorrect sentences, reflects the lower probability of the ungrammatical sentences as compared to the grammatical sentences (Coulson, King, & Kutus, 1998; but see Osterhout & Hagoort, 1999).

SUMMARY AND PREDICTIONS

The research reviewed above suggests that the statistical distribution of the inflectional suffixes within a paradigm can modulate isolated word recognition, so that words with higher information content are judged more slowly and lead to a larger amplitude brain response. Second, EEG research has shown that violations of a morphosyntactic constraint in sentence or phrasal context are registered as early as $0.3 \, s$ after the onset of a violation. The general aim of the present research is to determine whether the statistical distribution of inflectional suffixes in the German adjectival paradigm will modulate the cortical response to the adjective in phrasal context.

In phrasal context, information content might influence the process of recognising a word or its integration into the previous grammatical and semantic context. Previously, Vainio et al. (2008) have proposed that effects as late as the P600 might be taken as relatively late integrative effects, while ERP effects that are much earlier (e.g., $< 0.2 \, s$) might be taken as an influence on the word recognition process itself. On the other hand, it is clear from the grammatical violation literature that morphosyntactic violations can elicit LAN effects that begin approximately $0.3 \, s$ after the onset of a critical word or morpheme. Thus, we take $0.3 \, s$ as the lower bound for the beginning of integrative processing. To the extent that information content effects (including interactions) begin before $0.3 \, s$ this would be taken as evidence that it has likely influenced word recognition rather than integration. Effects that begin later are more likely to reflect phrasal integration.

In the experiment described below, these issues are investigated by comparing the responses of nonviolation prepositional phrases in strong and weak configurations. The phrases are presented using prepositions with different grammatical case requirements in order to test whether any observed differences between the strong and weak configurations depend on the specific grammatical case, and whether lexically-specific variables influence the response. In the experimental task, participants were presented with short prepositional phrases; half the phrases were grammatical, and half were ungrammatical. The task for participants was to judge whether the phrases were acceptable or not.

Several linking assumptions about the electrophysiological response should be made explicit (Barber & Kutas, 2007; Schall, 2004; Tanenhaus & Trueswell, 2004). The experiment reported here attempts to correlate electrophysiological activity recorded with EEG to a probability model for German morphosyntax. Although EEG has good temporal resolution relative to other noninvasive neurophysiological recording techniques, it is nevertheless only able to record volume currents that are aggregated over relatively large spatial and temporal scales relative to the activity of neurons or cortical columns (Nunez & Srinivasan, 2006). Our main assumption is that some of the electrical potentials present in patches of connected cortical areas responsible for morphosyntactic or lexical processing will have slow enough time dynamics in order to summate so that they can be observed with EEG.

Given these assumptions, we hypothesise that the amount of information conveyed by a word determines the integration of the word form into its context (e.g., forming a linked short-term memory representation with the previously established grammatical and semantic context). If the relative amount of morphosyntactic information conveyed by a word is higher, then the word will not be integrated as quickly because it will take longer to establish this linked memory representation. Our additional assumption is that the late ERP component magnitude will be greater when it takes longer to form this linked memory. If this assumption holds, the magnitude of the ERP response to the adjective should be larger in the strong form, because it carries more information. Our assumption is that creating a linked memory representation involves neural synchronisation within a given patch of cortex, and that this gives rise to amplitude differences observed in the average ERP. It should also lead to longer judgment times for the strong form to the extent that this delay is propagated to the motor system.

METHOD

Participants

Native right-handed German participants ($N = 20$) were recruited with posted advertisements from Radboud University in Nijmegen, a city near the east border of The Netherlands with Germany. The advertisements described a generic EEG experiment, and asked for native German-speaking participants. No participant reported problems with hearing, vision, or prior neurological injury.

Design, materials, and procedure

The design was intended to contrast the ERP response to the adjective in the strong versus the weak form of the phrases following the German inflectional paradigm (see Tables 1 and 3, and the earlier examples [1] and [2]. The main experimental variables

included the contrast between the two conditions strong and weak (*Dcln*), as well as the contrast between dative and accusative preposition (*Prep*), both variables within-subjects. The strong phrases were constructed without a determiner, while the weak phrases included a determiner. We used two different prepositions (*mit*, *ohne*) to provide two different cases (dative and accusative, respectively). To complete the phrases, we chose nouns corresponding to two genders (masculine, neuter) in singular form (see Table 3 for examples). In addition to the *Dcln* and *Prep* factors, the numerical covariates I_{w_x}, $I_{x|w}$, and D were calculated according the equations (7), (4), and (9), respectively. Multiple linear regression was used to orthogonalise these variables with respect to each other prior to modeling.

Along with the grammatically correct phrases analyzed here, an equal number of violation phrases were presented, consisting of a violation of declension (e.g., "mit dem *kleinem Boot"). These violation stimuli were similar to those presented to native speakers and second language learners in Davidson and Indefrey (2009). The original goal of the present design was to follow up on Davidson and Indefrey (2009) using a wider variety of lexical materials with native speaker participants. In this paper, we concentrate not on the violation-control contrast, but rather the control conditions primarily. This is done in order to test an assumption, namely whether the EEG sources underlying the violation response are also sensitive to nonviolation

TABLE 3
Examples of the experimental stimuli. The column *N* refers to the number of items presented in each experimental list

Declension	Preposition	Gender	Example	N
Weak	Dative	Masc	*mit dem großen Mann*	40
		Neu	*mit dem großen Haus*	40
	Accusative	Neu	*ohne das große Haus*	160
Strong	Dative	Masc	*mit großem Mann*	20
		Neu	*mit großem Haus*	60
	Accusative	Masc	*ohne großen Mann*	40
		Neu	*ohne großes Haus*	80

Filler category			Example	N
Violation-declension			*mit dem großem Mann*	40
			ohne das großes Haus	40
Control-other			*mit dem Haus*	40
			ohne den Mann	40
Violation-other			*mit großes Haus*	40
			ohne großem Mann	40
			mit dem großes Haus	40
			ohne das großem Haus	40
			mit große Haus	20
			mit große Mann	20
			mit das Haus	40
			ohne dem Mann	40
			ohne das große Mann	40
			mit großen Mann	80

Note: Haus (house), Mann (man).

morphological information, and if so, in what way. The violation stimuli are treated as fillers except for the purpose of selecting electrodes of interest (the violation response to the declension violation is used to define groups of electrodes for the later wavelet-regression analysis). It should be noted that for the nonviolation stimuli, the design is not completely crossed (or balanced) for the three factors of declension, preposition, and gender (see Table 3) because of our choice of the violation-control contrasts.

Also, note that the materials consisted of prepositional phrases, rather than full sentences. Although it would be preferable to employ full sentential stimuli, the reduced length of the word series allows for a greater number of trials.

The stimuli consisted of 587 adjective and noun pairs, chosen to capture most of the surface form frequency range, and were arranged into the phrases by a native German speaker. The adjectives and nouns were arranged into eight different lists. The stimuli were constructed by choosing adjective–noun pairs (for the masculine, and the neuter nouns, separately), and arbitrarily assigning groups of items to the conditions shown in Table 3. After creating the specific combinations of preposition, determiner (if necessary), adjective, and noun for each condition, the lists were randomised without any constraints separately for each participant. Table 3 shows the number of items for each combination of declension, preposition, and gender.

In the experimental task, the participants were asked to classify the prepositional phrases as grammatically acceptable or unacceptable following serial visual presentation at 0.6 s ISI/word (words were shown for duration of 0.3 s, with a blank period of 0.3 s between word onsets). The entire experiment consisted of one session in which a series of the phrases were presented as trials, with breaks approximately every 30 trials, or when requested by the participant. Participants were asked to relax or rest but were otherwise not provided a task during the breaks. Each trial began with a white fixation cross for 1 s, followed by the serial presentation of the phrase. At 0.5 s after the onset of the final word, another fixation cross was presented in yellow to indicate that a classification response was requested. The next trial began 2 s after the classification response.

Recordings and data analysis

EEG was recorded from 64 electrodes referenced to the left mastoid using battery-powered BrainVision BRAIN AMP Series amplifiers (Brain Products GmbH, München, Germany), later re-referenced to an average reference. Signals were sampled at 500 Hz, with a low-pass filter at 200 Hz and a high-pass filter with a time constant of 159 s during acquisition. Electrodes were applied to an equivalent inter-electrode distance Easy-Cap (Brain Products). Impedance levels were kept below 10 kΩ at the electrode-skin interface, with input impedance at the amplifiers at 10 MΩ (see Ferree, Luu, Russell, & Tucker, 2001). An additional electrode was placed below the left eye, referenced to an electrode above the eye, to record activity related to blinks or vertical eye movements. Lateral eye movement activity was recorded as the difference between channels at the left and right outer canthus.

The recorded EEG data were screened for eye movement, muscle, and other noise artifacts (artifact trials were excluded), filtered with a low-pass filter at 50 Hz (two-pass 6th-order Butterworth finite impulse response filter), and segmented into 1 s epochs consisting of 0.1 s before the onset of the critical word (CW) and 0.9 s following the CW. The resulting epochs were baselined with respect to the 0.1 s baseline interval and averaged according to experimental condition. Only trials with correct responses

were included in the analyses, and for the wavelet-regression analysis described below only grammatically correct phrases were analysed.

Two electrode groups were chosen to summarise the ERP response, based on the contrast between the declension violation and control conditions described earlier. There were two time windows for this contrast at 0.3 to 0.5 s and at 0.5 to 0.9 s, in line with the time windows typically chosen for the analysis of the LAN and P600 violation effects, respectively (see also Davidson & Indefrey, 2009). The analysis was conducted using a clustering and randomisation test (Maris & Oostenveld, 2007) over electrodes in these time windows. See Supplementary Section 1 for discussion of an alternative analysis based on analysis of variance (ANOVA). The goal of this analysis was to determine whether there was a larger amplitude response for the violation condition, and if so, to identify which electrodes were sensitive to this contrast. These electrodes were then used for the later wavelet-regression analysis of the control conditions. For the LAN-contrast, there was a significantly more-negative amplitude response for the violation condition compared to the control condition over eight left frontal electrodes (Supplementary Figure 1), $summedT = 20.12$, $p = .025$. For the P600 contrast, there was a significantly more positive amplitude response on 14 central-posterior channels for the violation condition compared to the control condition (Supplementary Figure 2), $summedT = 48.18$, $p < .001$. Based on these results, the left anterior group was comprised of the average of the electrodes near F7 (electrodes 8:9, 18:20, 33:34, and 39 on the Easy-Cap montage, see Supplementary Figure 3), and a posterior group comprised of the average of the electrodes surrounding Pz (electrodes 1, 4:7, 12:16, 26:27, and 29:30). Note that these sets of electrodes also correspond to the approximate locations at which the negative spatial mode of the LAN effect, and the positive spatial mode of the P600 violation effect have been observed in previous studies (e.g., Gunter et al., 2000; Koester et al., 2004).

A functional mixed effects regression model (Morris & Carroll, 2006) was used to estimate the effect of the regressors described in the Materials section. The functional regression is similar to mixed effect regression except that β coefficients are modeled as functions of time (see Davidson, 2009 for an overview and application to EEG data). Briefly, the single trials from an EEG experiment are transformed into wavelet coefficients using a discrete wavelet transform. Posterior coefficients for the regression are obtained using Markov chain Monte Carlo simulation, and the resulting coefficients are transformed back into the time domain. This approach also provides a Bayesian false discovery rate, calculated according to Morris, Brown, Herrick, Baggerly, and Coombes (2008). Posterior credibility intervals around the functional response are also derived. These can be interpreted as a confidence interval for the functional response that takes into account the dependencies between data points that arise in time series analyses. The credibility interval width is constrained so that $\alpha = 0.1$, and the minimum effect size was set at 0.1 μV, the recording amplifier resolution. In order to place the regression parameters on a common scale, factor regressors were coded as –0.5/0.5 and other predictors were standardised to a zero-mean, two standard deviation norm. For a similar, nonfunctional regression model approach, see Hauk, Davis, Ford, Pulvermuller, and Marslen-Wilson (2006). Note that the wavelet-based approach reported here includes only participants as a random effect. Although in principle both participants and items could be specified as (crossed) random effects with this approach, numerical instability prevented this in the present case.
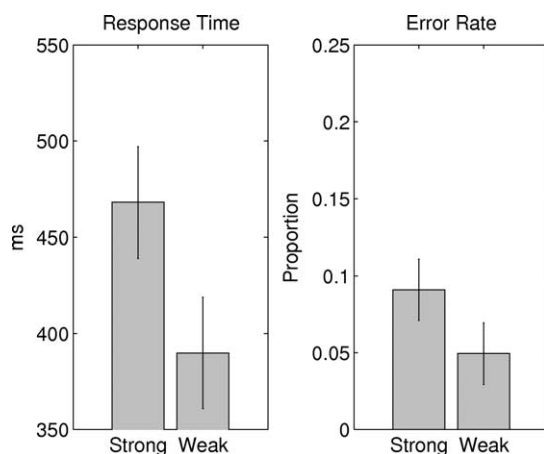
Figure 1. Average response times (left) and errors (right) for grammatical classification. Error bars indicate standard error of the mean over participants.

RESULTS

The behavioral responses suggested a slower response time for the strong form than the weak (see Figure 1). Classification times for the weak form were faster than the strong form by 87 ms (highest posterior density interval = 30, 136 ms). Subjects also made slightly fewer errors with the weak form (4%) versus the strong form (7%; $p = .001$). There were no main effects of preposition, nor an interaction between declension and preposition, for either response times or error rates.

In contrast to the behavioral results, the ERP results suggested a more complex pattern of results. The grand average ERPs for the strong and weak responses, for left frontal and posterior electrode groups, are shown in Figure 2, for the dative (*mit*) and accusative (*ohne*) prepositions.

The posterior electrode group shows a main effect of the strong–weak contrast for the accusative (beginning at approximately 0.4 s, see Figure 2d), but no effect is apparent for the dative. While there is some indication of a difference between strong and weak conditions for the frontal electrode group for the dative preposition at approximately 0.3 s (Figure 2a), the evidence is not strong.[1]

The functional regression analysis supported the above description for the main grammatical factors, and in addition showed statistical relationships with several of the information quantity regressors in several time windows (see Figure 3 for both the left frontal and the posterior electrode groups). These will be described below. In all of the plots, the highlighted time points are the mean values that exceed the minimum effect size and whose estimated probability exceeds the false discovery rate threshold. Additional analyses (conducted separately) including word length, simple form frequency, simple lemma frequency, and trial block showed no evidence of main effects or interactions with declension or preposition, so these factors will not be considered further. See Supplementary Section 2 for the analysis of the factor gender.

For the frontal group, the first plot (Ave) of Figure 3a shows the average response at the left frontal interest region to the weak condition, at the average level of the other

---

[1] Note that with a more restricted set of electrodes centered around the approximate location of F7, there is stronger evidence for this Dcln*Prep interaction, with the same functional form as shown in Figure 2a.
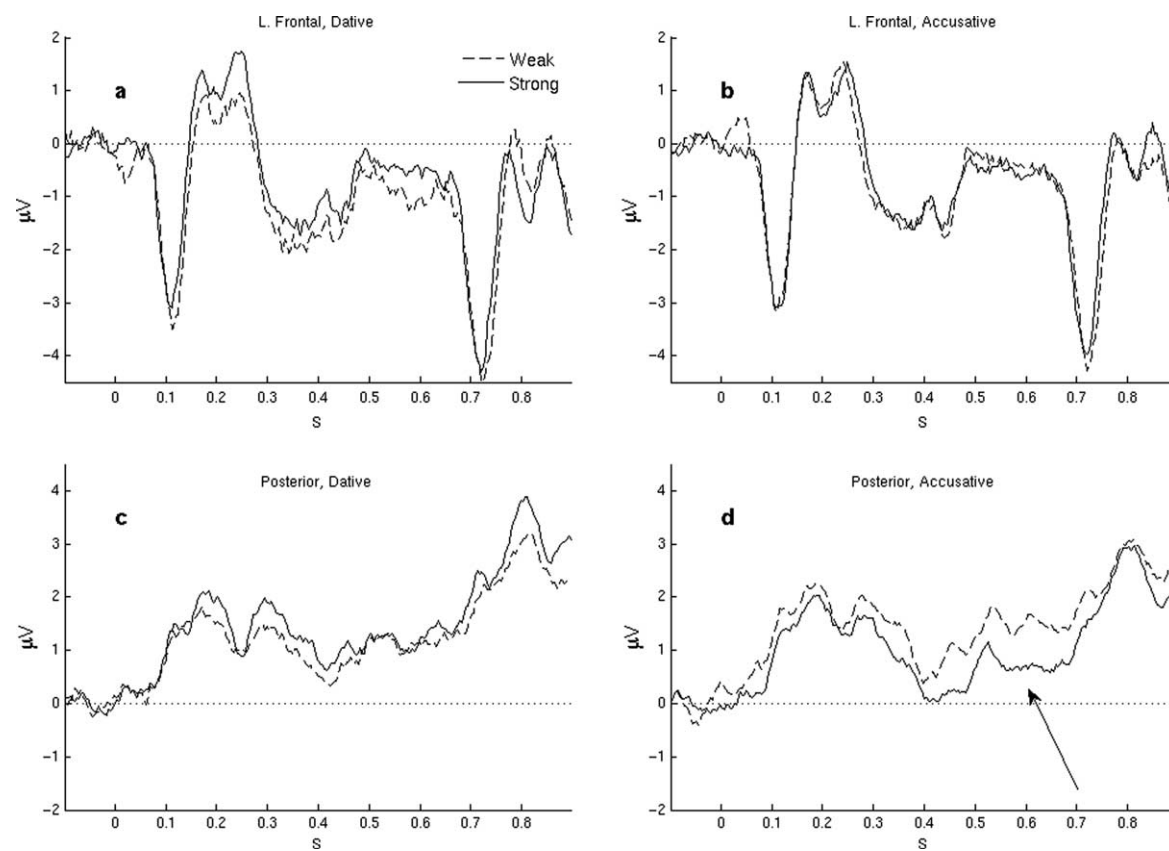
Figure 2.   Grand average event-related potential response to the strong and weak forms of the adjectives for the left frontal and posterior electrode groups, for the dative and accusative prepositions.
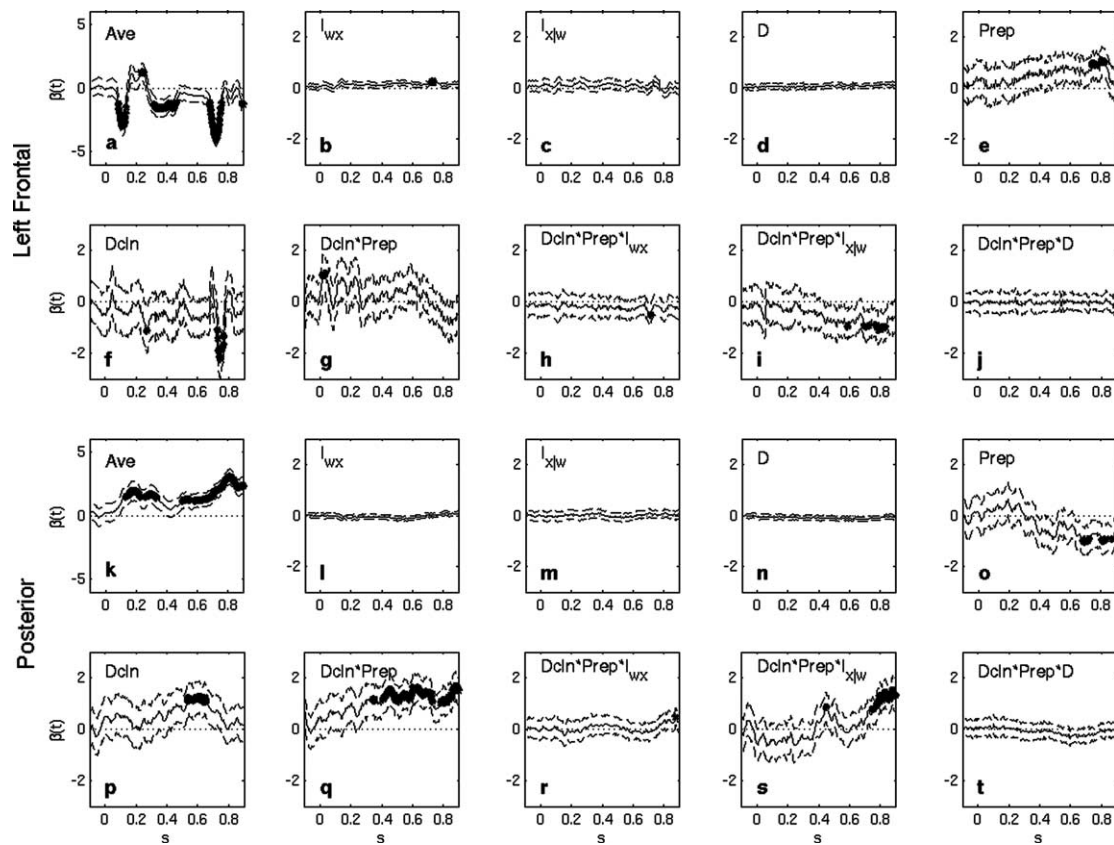
Figure 3. Coefficients (functional beta weights) for the left frontal (a–j) and posterior (k–t) electrode groups for the effects of the experimental factors declension (*Dcln*) and preposition (*Prep*), as well as interactions with several information quantities (see main text for description).
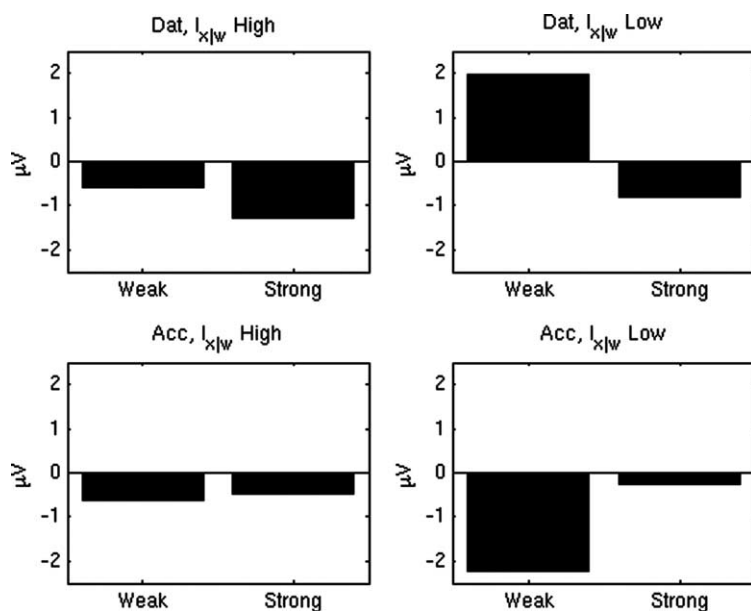
Figure 4. Average potentials for the interaction of declension (*Dcln*) and preposition (*Prep*), at high and low values of $I_{x|w}$, for the left frontal electrode group.

variables. The highlighted points are those that are likely to be greater or less than zero (i.e., the grand average response). There were no main effects for I_{w_x}, $I_{x|w}$, or D. There is a brief main effect of *Prep* at approximately 0.8 s, and of declension (*Dcln*) also at a late interval (0.75 s). There is no strong evidence of an interaction (*Dcln*Prep*) between declension and preposition on the frontal electrodes. Finally, the information measure $I_{x|w}$ modulated the interaction between *Dcln* and *Prep* in a late time window (Figure 3i). The interaction was more negative at approximately 0.6 to 0.8 s when the paradigm-specific information $I_{x|w}$ was greater. This modulates the nonsignificant late negative difference that is apparent in Figure 3g. Figure 4 shows the average ERP amplitudes for strong and weak declension in this late time window for the two cases, for high and low values of $I_{x|w}$ (based on a median split). The difference between strong and weak is similar for the dative and accusative case for high values of $I_{x|w}$. In contrast, for low values of $I_{x|w}$, the dative case shows the opposite pattern as the accusative case. For the dative case, there was a greater magnitude response for the weak declension compared to the strong but for the accusative, the weak declension led to a more negative response than the strong. In sum, for the frontal electrode group the interaction indicates that $I_{x|w}$ modulates the response to the declension contrast, depending on the type of preposition, relatively late within the time series.

On the posterior electrodes, there were significant main effects of preposition (Figure 3o) and declension (Figure 3p), as well as a similar interaction with the information quantity regressors (see Figure 3s) as was the case with the frontal response. In addition, there was a two-way interaction between declension and preposition (Figure 3q), indicating that for the accusative preposition, the strong declension had a more negative potential than the weak, starting at approximately 0.4 s and remaining until approximately the end of the observation interval (see also Figure 2). The brief main effects of preposition and declension should be seen in light of this interaction. The paradigm-specific information content $I_{x|w}$ modulated the
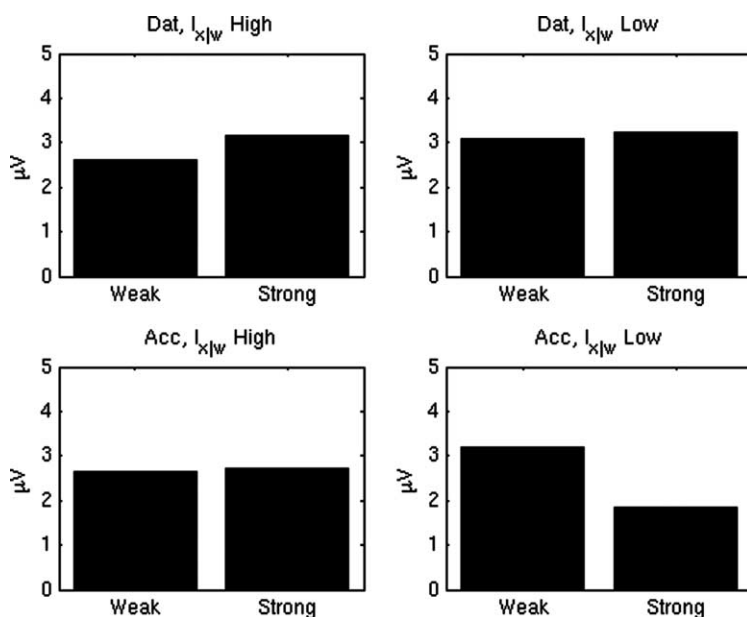
Figure 5. Average potentials for the interaction of declension (*Dcln*) and preposition (*Prep*), at high and low values of $I_{x|w}$, for the posterior electrode group.

interaction between declension and preposition (Figure 3s) in a sustained late time window, starting at approximately $0.8\,s$.

Figure 5 shows the average ERP amplitudes for strong and weak declension in this late time window for the two cases, for high and low values of $I_{x|w}$. The difference between strong and weak is similar for the dative case for both high and low values of $I_{x|w}$ (again, based on a median split). In contrast, for the accusative case there was no difference between strong and weak for high values of $I_{x|w}$, but a greater magnitude response for the weak declension compared to the strong for low values of $I_{x|w}$. As with the frontal response, this interaction indicates that $I_{x|w}$ modulates the response to the declension contrast, depending on the type of preposition, but only relatively late within the time series.

There were also several other brief higher-order interactions (e.g., Figure 3g,h,r), but because they were not sustained they will not be considered here further.

DISCUSSION

The experiment reported here has shown a differential ERP response to the strong and weak declension of the German adjective, presented in a short prepositional phrase via rapid serial presentation. Also, participants were slower and (slightly) less accurate to give judgments for phrases presented in the strong form, compared to the weak. In addition, the adjective-specific information content of the inflection (e.g., $I_{x|w}$) affected the ERP response to the adjective, but at a relatively late time interval (i.e., at approximately $0.8\,s$). This provides positive evidence that lexically-specific variables can modulate the ERP response in phrasal context.

The behavioral results are largely congruent with previous studies of the influence of sentential or phrasal context on lexical decision times. Previous experiments have shown that readers and listeners can use phrasal context to aid lexical decision-making

(Bölte & Connine, 2004; Gurjanov, Lukatela, Lukatela, Savić, & Turvey, 1985; Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985). The present results could be interpreted to reflect a similar process. When participants encounter a German adjective within an attributive adjectival phrase that also includes a determiner, the determiner provides additional grammatical context for the processing of the adjective and the noun to appear after the adjective. In addition, the grammatical information is presented over a longer time period with the weak form of the phrases, allowing comprehenders more time to integrate the information from each additional word before encountering the end of the phrase. These factors would explain why participants were faster to judge the weak forms of the phrases than the strong.

In the behavioral task, participants responded with the meta-linguistic judgment at the end of the phrase, so it is not clear when the differential effect of declension might have arisen. The judgment was cued at $0.5\,s$ after the end of the phrase, meaning that short-lived effect might have been missed using the judgment procedure employed here. Future studies using acceptability judgments during the presentation of the phrases themselves or reading time studies (e.g., eye-tracking) might be informative about the timing of the effects of declension.

Nevertheless, the ERP results do offer some evidence about when the effects might arise. The ERP contrasts revealed that with the accusative (but not dative) prepositions for the posterior electrode group, there was a differential response in a time window beginning at approximately $0.4\,s$. The timing of this response difference would suggest that this activity could reflect grammatical or semantic integration, as previous ERP studies have shown that the effect of grammatical or semantic context most often is apparent at this time (for review see Kutas, Van Petten, & Kluender, 2006). Beyond this, it is not clear why the two different prepositions would lead to a differential response. Although the accusative preposition *ohne* ("without") also conveys a difference in semantic meaning (negation) compared to the dative *mit* ("with"), this does not straightforwardly lead to the prediction of the difference that was observed.

Another potential explanation for the differential response might be a difference in featural information. If we follow Schlenker's (1999) general idea of featural information accumulating from left to right then we can ask what new information the adjective form provides in the series of words. In the task the prepositions were always present, so case information was present before the determiner or adjective. In the *weak* conditions, the determiners provide the following new information:

[5]   mit (dative)   dem $(-F, -Pl)$

That is, gender is incompletely specified (masc or neuter), whereas number is specified.

[6]   ohne (acc)   den $(+M, -Pl)$
[7]   ohne (acc)   das $(+M -F, -Pl)$

In the accusative case, gender is fully specified, as well as number. In the three weak conditions, the following adjective with suffix *–en* provides no new information.

In the *strong* conditions, the adjectives provide the following new information:

[8]   mit (dative)   –m $(-F, -Pl)$

That is, in the dative case, gender is incompletely specified (masc or neuter), and number is specified.

[9]   ohne (acc)   –n $(+M, -Pl)$
[10]   ohne (acc)   –s $(+M -F, -Pl)$

In the accusative case, gender is fully specified, as well as number.

Under this feature account, the weak adjective form provides no new information for dative or accusative. The strong form provides new information: plural for both dative and accusative, but fully specified gender information only for accusative (for dative, only underspecified gender information). It may be the case that the posterior electrode group picks up this difference between the amount of gender information provided by strong forms in dative and accusative contexts from $0.4\,s$ onwards. This difference may then be further modulated by adjective-specific information content of this suffix relative to all forms of the particular adjective stem at later stages (after $0.8\,s$). In future work, it would be advisable to use a greater variety of prepositions to determine whether the interaction is due to a global property of case, or lexically-specific properties, or both. It would also be useful to obtain a source reconstruction for this response difference to determine whether it results from activity within distinct cortical areas (e.g., frontal and temporal areas), or alternatively, activity within a similar region but with a difference in orientation or geometry. This would help constrain the interpretation of the response difference to the extent that semantic or grammatical effects are differentiated by anatomical location.

The adjective-specific information content modulated the interaction of case and declension. This result supports the contention of earlier work, conducted using lexical decision tasks, that the statistical organisation of words within a morphological paradigm has an influence on word processing (e.g., Milin et al., 2008). In addition the results show that these effects can be seen in adjectives, as some previous experiments have observed that adjectives are not always sensitive to these statistical variables (Kostić & Katz, 1987). Perhaps most importantly, however, the results presented here extend this finding to phrasal contexts, and provide some evidence about the timing of an electrophysiological correlate of this factor. The relatively late time interval for the interaction is consistent with earlier reports from studies employing eye movement recordings (Hyönä et al., 2002; Vainio et al., 2008). These studies have shown that morphological complexity variables seem to exert an influence relatively late within the parsing and interpretation process during the processing of sentences. In the present study, although the grammatical context factors influenced the ERP response magnitude beginning at approximately $0.4\,s$, the lexically-specific information modulated the response only late within the response interval.

Although these ERP effects were present in grammatical phrases, it is worthwhile considering in what ways they might be similar to previous ERP results using either ungrammatical or noncanonical phrases. In earlier studies of violation effects (see Introduction), two or three different stages of processing have been proposed. In the earliest stage, lexical access occurs, during which morphosyntactic or semantic features as accessed. During later stages, context integration, and in the case of input with a grammatical error, re-analysis takes place. If one assumes that these stages correctly characterise the parsing process, then the current results suggest that the influence of statistical factors related to the inflectional paradigm operate primarily on the later integration stage. Because the interaction was observed relatively late, it seems more likely that integrative rather than lexical access processes were involved. Also, because the interaction was observed in grammatically correct phrases, it seems unlikely that a process of re-analysis would be involved. Earlier reports have shown LAN and P600 effects for noncanonical compared to canonical phrasal order in German (e.g., Rösler et al., 1998), and it appears that this type of effect would be more similar to the present study. In studies of this response in German, the first manifestation is observed at the determiner of the first noun phrase within the sentence as a greater amplitude response to noncanonical phrasal order. In this sense,

the effect can be seen as an interaction of inflectional morphology of the determiner with the grammatical structure of the construction. There are important differences, however. First, the canonicity effect is first observed within the time window of the LAN, while the interaction here was first seen within the time window of the P600. Second, the canonicity effect is seen in response to a determiner, while the interaction here was present for inflected adjectives. Perhaps one reason for the difference in timing could be due to this difference in word class, or possibly greater complexity associated with identifying the suffix given the adjective, compared to the determiner. In any case, the present results may imply that modulation of late components may reflect the statistical information content of the inflected forms in grammatical context. Future work might explore how probabilistic relationships at the phrasal level, in addition to the morphosyntactic level, influence the electrophysiological response.

One important caveat for this interpretation is that EEG, like all noninvasive physiological measures, is selective for certain types of brain activity (as described in the Introduction). It cannot be concluded, for example, that an earlier modulation of the response is not present in cortex. Rather it was not apparent for the present recording setup and design. In the absence of any evidence that the modulation is earlier, however, the most parsimonious conclusion is that the influence only occurs relatively late in the response interval. Future work might examine both EEG and eye-tracking performance using the same materials, to determine whether there is item-covariance between the effects seen in EEG and eye-tracking. Also, it is important to consider that the responses reported here are restricted to the left frontal and posterior electrode groups. It may be possible in future studies to extend the wavelet-regression technique employed here to include spatial topography as well as time, or ideally, to source waveforms (e.g., see Solomyak & Marantz, 2010). A final caveat is that only certain parts of the German adjective declension were examined here. Future studies might obtain a better picture of the response to the adjectival declension contrasts by sampling adjectives that occur in a greater variety of cases and genders. The present results serve as an initial estimate of the response for the paradigm.

In common with some of the lexical decision experiments reviewed in the Introduction, the present study employed a task in which participants provided a meta-linguistic judgment about the acceptability of the phrases. It might be the case that the ERP patterns we observed were influenced by the judgment task, and it is worth considering whether the need for a judgment may have influenced the late responses that we observed. It seems possible that the late interaction effects seen in the present study are analogous to a late verification phase of the P600 violation effects seen previously in the literature (see Introduction). This might have been the case if, as suggested by an anonymous reviewer, the presence of violation stimuli in the experiment, and/or the need to make acceptability judgments, induced participants to employ the same verification mechanisms on the control trials as on the violation trials. This cannot be excluded in the present design. Future studies might employ a task that does not involve meta-linguistic judgments, and which does not include violation stimuli, such as passive reading or listening. This would take advantage of the property of EEG measurements, which do not require that a specific behavioral task is used in order to measure ERP components.

At a very general level, the results demonstrate that ERP amplitude differences can be observed without (necessarily) contrasting violation and control stimuli, which remains one of the most popular methods for using electrophysiology to study sentence comprehension. The contrast between strong and weak declension led to a

differential response within grammatically licensed phrases, and this result was apparent in the presence of several other regressors for lexical factors. These functional regression results also add further support to the general approach of using linguistically conditioned statistical variation as a means of investigating the neural correlates of language processing, and ultimately the function of the networks that are responsible for linguistic comprehension. Although both approaches have been taken (use of nonviolation paradigms, the influence of lexical factors) previously in the EEG sentence processing literature, the intended contribution of the present experiment was to illustrate how linguistically conditioned morphosyntactic variation might be reflected in the ERP response to words in phrasal context, where the variation is conditioned by the phrasal context. While the results presented here are preliminary in many respects due to design and task limitations, the general approach may prove useful in further investigations of declension and other morphosyntactic processes, especially those that attempt to link the results from behavioral and electrophysiological methods.

REFERENCES

Baayen, R. H., McQueen, J., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 355–390). Berlin: Mouton de Gruyter.

Barber, H. A., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17, 137–153.

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, 53, 98–123. doi:10.1016/j.brainresrev.2006.07.002.

Bierwisch, M. (1967). Syntactic features in morphology: General problems of so-called pronominal inflection in German. In *To honor Roman Jakobson* (pp. 239–270). The Hague: Mouton.

Blevins, J. P. (1995). Syncretism and paradigmatic opposition. *Linguistics and Philosophy*, 18, 113–152.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93.

Bölte, J., & Connine, C. M. (2004). Grammatical gender in spoken word recognition in German. *Perception & Psychophysics*, 66, 1018–1032.

Cahill, L., & Gazdar, G. (1997). The inflectional phonology of German adjectives, determiners, and pronouns. *Linguistics*, 35(2), 211–245.

Clahsen, H., Eisenbeiss, S., Hadler, M., & Sonnenstuhl, I. (2001). The mental representation of inflected words: An experimental study of adjectives and verbs in German. *Language*, 77, 510–543.

Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21–58.

Davidson, D. J. (2009). Functional mixed-effect models for electrophysiological responses. *Neirofiziologiyal Neurophysiology*, 41, 79–87.

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience*, 21, 433–446.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language*, 45, 200–224.

DWDS-Kerncorpus. (2008). *Berlin-Brandenburgische Akademie der Wissen-schaften* [Berlin Brandenburg Academy of Science]. Berlin. Available from: www.dwds.de.

Faussart, C., Jakubowitz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Revista di Linguistica*, 11, 357–374.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, 112, 536–544.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory. Cognition*, 22, 1219–1248.

Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In C. Fellbaum (Ed.), *Collocations and idioms: Linguistic, lexicographic, and computational aspects.* (pp. 23–41). London: Continuum Press.

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

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 692–701.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19,* 55–71.

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

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

Hauk, O., Davis, M. H., Ford, M., Pulvermuller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage, 30,* 1383–1400.

Hyönä, J., Vainio, S., & Laine, M. (2002). A morphological effect obtains for isolated words but not for words in sentence context. *European Journal of Cognitive Psychology, 14,* 417–433.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP study. *Journal of Psycholinguistic Research, 31,* 165–193.

Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience, 16,* 1647–1668.

Kostić, A. (1991). Informational approach to processing inflected morphology: Standard data reconsidered. *Psychological Research, 53,* 62–70.

Kostić, A. (1995). Informational load constraints on processing inflected morphology. In L. B. Feldman (Ed.), *Morphological aspects of language processing.* Hillsdale, NJ: Lawrence Erlbaum.

Kostić, A., & Katz, L. (1987). Processing differences between nouns, adjectives, and verbs. *Psychological Research, 49,* 229–236.

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

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of psycholinguistics* (2nd ed, pp. 83–143). New York: Elsevier.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177.

Lukatela, G., Kostić, A., Feldman, L., & Turvey, M. T. (1983). Grammatical priming of inflected nouns. *Memory & Cognition, 11,* 59–63.

Lukatela, G., Kostić, A., Todorovi, D., Carello, C., & Turvey, M. T. (1987). Type and number of violations and the grammatical congruency effect in lexical decision. *Psychological Research, 49,* 37–43.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190.

Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2008). Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 214–278). Oxford: Oxford University Press.

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research, 1228,* 161–176.

Morris, J. S., Brown, P. J., Herrick, R. C., Baggerly, K. A., & Coombes, K. R. (2008). Bayesian analysis of mass spectrometry proteomics data using wavelet based functional mixed models. *Biometrics, 64,* 479–489.

Morris, J. S., & Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, B, 68*(2), 179–199.

Moscoso del Prado Martín, F., Kosti, A., & Baayen, R. H. (2004). Putting the bits together: An information-theoretical perspective on morphological processing. *Cognition, 94,* 1–18.

Münte, T. F., & Heinze, H. J. (1994). ERP negativities during syntactic processing of written words. In H. J. Heinze, T. F. Münte, & G. R. Mangun (Eds.), *Cognitive electrophysiology: Basic and clinical applications* (pp. 211–238). Boston: Birkhüser.

Münte, T. F., Heinze, H. J., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience, 5,* 335–344.

Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruities in words and pseudo-words. *Journal of Cognitive Neuroscience, 9,* 300–311.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164,* 81–94.

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysiology of EEG* (2nd ed). Oxford: Oxford University Press.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, *14*, 1–14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739–773.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.

Penke, M., Janssen, U., & Eisenbeiss, S. (2004). Psycholinguistic evidence for the underspecification of morphosyntactic features. *Brain and Language*, *90*, 423–433.

Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition*, *91*, B35–B45.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, *38*, 150–176.

Schall, J. D. (2004). On building a bridge between brain and behavior. *Annual Review of Psychology*, *55*, 23–50.

Schlenker, P. (1999). La flexion de l'adjectif en allemand: la morphologie de haut en bas [The inflection of adjectives in German: Morphology from top to bottom]. *Recherches linguistiques de Vincennes*, *28*, 115–132.

Silva-Pereyra, J., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research*, *1185*, 201–211.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*, 2042–2057.

Tanenhaus, M. K., & Trueswell, M. K. (2004). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to processing world-situated language: Bridging the product and action traditions* (pp. 3–38). Cambridge, MA: MIT Press.

Vainio, S., Hyönä, J., & Pajunen, A. (2008). Processing modifier-head agreement in reading: Evidence for a delayed effect of agreement. *Memory & Cognition*, *36*, 329–340.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.

Wunderlich, D. (1997). Der unterspezifizierte Artikel [The underspecified article]. In M. S. U. & K.-H. R. Christa Drscheid (Eds.), *Sprache im Fokus* (pp. 47–55). Tübingen: Niemeyer.

Zwicky, A. M. (1986). German adjective agreement in GPSG. *Linguistics*, *24*, 957–990.