The role of feature-number and feature-type in processing Hindi verb agreement violations

Andrew Nevins[1]
Brian Dillon[2]
Shiti Malhotra[2]
Colin Phillips[2,3]

[1]*Department of Linguistics, Harvard University,* [2]*Department of Linguistics, University of Maryland, College Park* [3]*Neuroscience and Cognitive Science Program, University of Maryland, College Park*

Total number of pages: 41
Number of tables: 7 (incl. appendix)
Number of figures: 2

Address for correspondence:

Andrew Nevins
Department of Linguistics
Harvard University
317 Boylston Hall
Cambridge, MA 02138

nevins@fas.harvard.edu
(617) 495-8107 (ph.)
(617) 496-4447 (fax)

Preparation of this paper was supported by grants to CP from the National Science Foundation (BCS-0196004) and the Human Frontier Science Program (RGY-0134) and by a Semester Research Award from the University of Maryland. We are very grateful to Shradha Upadhyay for assistance in constructing materials and recruiting participants and to Alison Austin for assistance with data collection and analysis.

Abstract

This article presents study of Hindi that investigates whether responses to syntactic agreement violations vary as a function of the type and number of incorrect agreement features, using both electrophysiological (ERP) and behavioral measures. Hindi is well suited to investigation of this issue, since verbs in Hindi mark agreement with the person, number, and gender features of the nominative subject noun phrase. Evoked responses were recorded for visually presented verbs appearing at the end of a sentence-initial adverbial clause, comparing responses in a grammatically correct condition with four grammatically incorrect conditions that mismatched the correct agreement on different dimensions (Gender, Number, Gender/Number, Person/Gender). A P600 response was elicited in all grammatically incorrect conditions. No amplitude differences were found among the Gender, Number, and combined Gender/Number violations. This suggests that the feature distance between observed and expected word forms at the morphosyntactic level does not impact ERP responses, contrasting with findings at the semantic level (Federmeier & Kutas, 1999) and in auditory processing (Paavilainen, Valppu, & Näätänen, 2001), and suggests that the P600 response to agreement violations is not additive based on the number of mismatching features and does not reflect top-down, predictive mechanisms. A significantly larger P600 response was elicited by the combined Person/Gender violation, and two different violations involving the Person feature were judged as more severe and recognized more quickly in the behavioral studies. This effect is attributed to the greater salience of the Person feature at multiple levels of representation.

Theme: Neural basis of behavior

Topic: Cognition

Keywords: Event-related potentials, sentence processing, agreement, P600, Hindi

1. Introduction

1.1 Agreement Processing Mechanisms

In this article we examine the processing of rich verbal agreement morphology in Hindi, examining the contribution of bottom-up and top-down mechanisms in agreement processing. This has relevance for broad questions about the uniformity of processing mechanisms across semantics, syntax, and phonetics/phonology. It also addresses questions about the grain-size of morphosyntactic analysis.

Agreement is a widespread phenomenon in natural language, marking concord between noun phrases (NPs) and verbs as in *he runs* vs. *they run*, or between nouns and determiners or adjectives as in *this house* vs. *these houses*. There is substantial cross-language variation in the range of relations that are marked by agreement and in the morphological properties that are marked by agreement, including person, number, gender, definiteness and case. There is a large linguistic literature on the nature of agreement systems cross-linguistically (Corbett, 2006; Wechsler & Zlatic, 2003). Many psycholinguistic studies have examined agreement processing, particularly through studies of *attraction errors* in agreement production (Bock & Miller, 1991; Eberhard, Cutting, & Bock, 2005) and comprehension (Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey, & Bock, 1999; Solomon & Pearlmutter, submitted). These studies are complemented by a growing body of research that tracks the processing of agreement violations using electrophysiological measures (Coulson, King, & Kutas, 1998; Hagoort, Brown, & Groothusen, 1993; Kaan, 2002; Osterhout & Mobley, 1995). It is now clear that speakers are highly sensitive to agreement relations and that agreement errors are typically detected reliably and within a few hundred milliseconds. There is also some evidence that different agreement features may be processed differently, based primarily on detailed timing information (Barber & Carreiras, 2005; de Vincenzi, 1999).

Despite the rich body of evidence on agreement processing, less is known about the specific processes that lead to the checking of correct agreement between two elements in a sentence. Central to the agreement checking must be a process of *comparison* that evaluates the compatibility between a pair of representations. However, it is less clear what representations are compared and whether the evaluation process involves a unitary mechanism or a series of partially independent mechanisms. Let us consider these two questions and related background in turn, focusing on the case of a verb that agrees with a subject NP that precedes it.

First, it is clear that subject-verb agreement must be checked by evaluating the match between the agreement morphology of the verb and the relevant properties of the subject NP, but this could be achieved in a few different ways. Under a strongly *bottom-up* approach to agreement, agreement processing in a subject-verb sequence does not begin until the agreement morphemes are identified upon presentation of the verb. Processing of the verb initiates a search for a subject NP, which must be analyzed to determine whether it bears appropriately matching features. Alternatively, under an approach to agreement checking that includes a *top-down* component, the processing of the features that must be matched can start before presentation of the verb. Processing of the subject NP may already trigger identification of the features that are relevant for agreement. Furthermore, the position of the agreeing verb may be predictively created and its agreement features pre-built. If the morphology of the language is suitably predictable, then even the specific forms of the verb agreement might be anticipated. Under this approach, presentation of the verb does not entail a search for a suitable subject NP to agree with, and requires merely that the features of the verb be matched to the features that have been predictively built.

There has been much interest in the scope of top-down or predictive processes in language understanding, spanning older and more recent theoretical and computational models (Crocker, 1996; Elman, 1991; Frazier & Fodor, 1978; Gibson, 1998; Hale, 2003; Kimball, 1975; Levy, 2006). A number of recent studies present evidence for predictive mechanisms, based on studies using reading-time measures (Chen, Gibson, & Wolf, 2005; Staub & Clifton, 2006) and event-related brain potentials (Lau, Stroud, Plesch, & Phillips, 2006), and based on anticipatory looks in head-mounted eye-tracking studies

(Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003). It has also been argued that evidence for incremental processing of verb-final languages such as German and Japanese entails predictive structure building operations (e.g., Bader & Lasser, 1994; Aoshima, Yoshida, & Phillips, submitted).

Despite the interest in predictive processes in language comprehension, it remains unclear whether agreement processing involves bottom-up processes that search for an agreement controller or whether it involves comparison of an incoming word with a predictively created morphological template. The pervasiveness of agreement attraction errors in production and comprehension of sequences like *the key to the cabinets were* ... might suggest a bottom-up search process that causes a verb to sometimes incorrectly agree with a nearby NP. However, this interpretation is challenged by the finding that rates of attraction errors are best predicted by the closeness of the 'attracting' NP to the correct subject NP, and not by the closeness of the attracting NP to the verb (Bock & Cutting, 1992; Franck, Vigliocco, & Nicol, 2002; Vigliocco & Nicol, 1998). This is argued to show that the processor incorrectly encodes an NP like *the key to the cabinets* as plural, and not to imply that the processor incorrectly forms an agreement relation between the attracting NP and the main verb. In contrast, some evidence for predictive processing of morphology comes from sequential lexical decision studies in Serbo-Croatian by Lukatela and colleagues that show that agreeing verbs are recognized more quickly when preceded by an appropriately agreeing pronoun (Lukatela, Moraca, Stojnov et al., 1982) and that case-marked nouns are recognized faster when preceded by correspondingly case-marked adjectives (Gurjanov, Lukatela, Moskovljevic, Savic, & Turvey, 1985). Additionally, Solomon and Pearlmutter have argued for a partially top-down *forward-tracking* view of agreement processing, based on evidence from number-frequency effects in the processing of uninflected modal verbs like *would* and *must* in English (Solomon & Pearlmutter, submitted). Thus, one goal of the current study is to use an alternative measure to investigate the contribution of top-down (predictive) or bottom-up mechanisms in processing agreement morphology.

A second question of interest is whether or not the agreement checking mechanism is a unitary process, in other words whether it consists of one matching process or multiple matching processes that evaluate different agreement features. It is common for agreement to track multiple features of the controller, such as person, number, gender, or case, and these are commonly encoded using separate agreement affixes. It is possible that the agreement features of a verb are identified as a group during the process of word recognition and then compared as a group to the features of the subject NP (whether in a bottom-up search process or a top-down matching process). In this case the processor may simply determine that the agreement features match or do not match, and there may be no advantage for a set of features that partially matches (e.g., by mismatching in one feature only), relative to a set of features that fully mismatch. Alternatively, each agreement feature may be compared separately by partially independent sub-processes, such that it is possible for some sub-processes to find a match while other sub-processes find a mismatch. Under this scenario, the processor may distinguish between partial and full mismatches in agreement features, since they result in different error signals from the distributed comparison process.

There are few previous findings that directly address this issue, since most studies of agreement processing have not directly compared the processing of different agreement features, or the processing of partial vs. full mismatches. Nevertheless, there are some relevant previous findings, which appear to point to different conclusions. First, in a continuous lexical decision task Lukatela and colleagues tested sequences in which a possessive pronoun was followed by a noun that matched or mismatched in case, number, or gender, or some combination of these, and found that although there was a robust grammatical congruency effect on decision times, it was not affected by the number or type of features that mismatched (Lukatela, Kostic, Todorovic, Carello, & Turvey, 1987). Person features were not manipulated in this study, since the study focused on agreement within noun phrases. The results of this study suggest that agreement features are evaluated jointly and that the processor cares only about whether or not there is a perfect match. However, some other research suggests differences in the processing of number and gender features that may indicate distinct processing mechanisms. Cross-modal priming evidence from Italian pronoun resolution suggests that number cues are more rapidly used than gender cues in disambiguating the antecedent of a pronoun (de Vincenzi, 1999). ERP results on the processing of Spanish indicate that the late positivities elicited by agreement violations exhibit a later peak or a longer duration for gender violations than for number violations (Barber & Carreiras, 2005). In both cases, this difference is attributed to a representational difference between gender, an inherent property of nouns, and number, an independent feature. This difference may entail different recognition and reanalysis processes. However, it should be noted that these findings were based on agreement marking on nominals, and it is not clear whether they should generalize to agreement marking on verbs, since gender is not an inherent property of verbs. Thus, existing findings are inconclusive on the existence of a unified or distributed evaluation mechanism for different agreement features.

A related question is whether all agreement features are analyzed in the same manner, or whether an agreement mismatch in one feature (e.g. person) is analyzed as a more severe deviation than a mismatch in another feature (e.g. number).

It is possible to investigate these questions about the nature of agreement processing by testing the electrophysiological consequences of individual vs. multiple agreement feature violations in Hindi, a language with rich verb agreement morphology. Our approach builds upon previous electrophysiological studies that have shown gradient effects of anomaly in semantic or phonetic processing, and on previous studies on the processing of agreement violations.

1.2 Gradient ERP Responses to Linguistic Anomalies

The use of event-related brain potentials (ERPs) has led to the documentation of a series of temporally and topographically distinct response components that are elicited by different kinds of unexpected linguistic material. Words that are syntactically appropriate but semantically inappropriate characteristically elicit a central negativity known as the

N400 (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000). Infrequent oddball syllables that interrupt an otherwise uniform train of sounds elicit a response component with a 150-300 ms latency known as the mismatch negativity (MMN: Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). Words that are morphologically or syntactically incorrect elicit a late positivity known as the P600 (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), and in some cases also an (early) left anterior negativity ((E)LAN: Coulson, King, & Kutas, 1998; Friederici et al., 1993; Hagoort, Wassenaar, & Brown, 2003; Neville, Nicol, Barss, Forster, & Garrett, 1991). Since the initial discoveries of these response components there has been substantial interest in the goal of understanding what specific factors control the timing and amplitude of these effects. This interest has been guided by the assumption that an improved understanding of this variation will lead to more precise models of the representations and processes that underlie the ERP effects.

Quantitative variation in ERP responses elicited by unexpected linguistic material has been investigated in detail in the domains of semantic interpretation and auditory processing, and these therefore serve as a useful point of comparison to the morphosyntactic phenomena investigated here. Previous research on semantics and auditory perception has shown that quantitative variation in ERP response amplitudes is associated with at least two different dimensions of the relation between expected and unexpected material.

Mismatches between expected and unexpected/incongruous linguistic items may be quantified in terms of relative probabilities of occurrence, or in terms of the degree of similarity (or 'matching') in properties between the expected and unexpected items. In studies of semantic processing much evidence indicates that N400 amplitude varies as a function of the probability that the target word should appear given its preceding context. For example, following the context *She was stung by a* ... words that are implausible or anomalous, such as *fly* or *dog*, are clearly less probable than the word *bee*, and thus elicit larger N400 responses (Kutas & Hillyard, 1980). Furthermore, a final word that is plausible but less probable, such as *wasp*, also elicits a larger N400 response than the most probable completion. Thus, N400 amplitude is considered to inversely correlate with the *cloze probability* of a syntactically appropriate content word (Kutas & Hillyard, 1984; van Petten & Kutas, 1990, 1991). In the domain of syntax there is some evidence for similar variation in P600 amplitude as a function of the probability of different types of arguments following a verb (Osterhout, Holcomb, & Swinney, 1994).

However, recent evidence suggests that N400 amplitude may also vary as a function of the degree of semantic overlap, as quantified by a feature-space, between a target word and the expected word, independent of cloze probability. A study by Federmeier & Kutas varied the semantic properties of sentence-final words in highly constraining contexts in an attempt to investigate the role of memory organization in linguistic processing (Federmeier & Kutas, 1999). They found that pairs of words that were entirely implausible and improbable but varied in their semantic feature overlap with the most expected word elicited different amplitude N400 responses, with a reduced N400 in cases of greater semantic overlap. In (1), for example, the word *pines* elicited a smaller N400 than *tulips*, relative to the expected final word, *palms*. Federmeier & Kutas suggest that this feature-distance effect reflects prediction of the expected word based on context, which in turn leads to priming of the semantic features of words that are inappropriate yet related to the expected word. Therefore, the finding of gradient N400 responses to semantically anomalous words that are equally improbable provides evidence for a top-down predictive mechanism that activates highly probable words before they appear in the input.

(1) They wanted to make the hotel look more like a tropical resort. So, along the driveway they planted rows of {palms, pines, tulips}.

Evidence for a similar feature-overlap advantage has been found in ERP studies of auditory processing. For example, Levänen and colleagues showed in an auditory mismatch paradigm that when the deviant sound mismatched the standard sound along two dimensions (frequency and duration) a magnetic mismatch field (MMF) was elicited that was roughly the sum of the individual MMFs elicited by either mismatched dimension individually (Levänen, Hari, McEvoy, & Sams, 1993). These findings were extended in a subsequent study in which stimuli included two- and three-feature deviants, manipulating the dimensions of frequency, intensity, and stimulus-onset asynchrony (Paavilainen, Valppu, & Näätänen, 2001). This study found that MMN amplitudes increased with greater featural distance between the standard and deviant sounds, although amplitudes were additive for 2-feature mismatches and less than additive for 3-feature mismatches. These findings are compatible with a standard account of the auditory mismatch response as a reflection of a process of comparison between a deviant stimulus and a memory trace of the standard sound (Näätänen, 1992).

Thus we find evidence in at least two domains for feature priming, understood as a process in which the features of a likely upcoming item are primed. By hypothesis, priming effects can be identified based upon reduced amplitude of the MMN or N400, which implies reduced neural activity, and hence reduced processing requirements. We next turn to the ERP components characteristically associated with agreement processing.

Many previous studies have documented the characteristic ERP responses associated with agreement violations. A common finding is that verb inflection errors involving agreement or tense elicit a biphasic response consisting of an anterior negativity that is sometimes left-lateralized (LAN) followed by P600 response with a broad posterior scalp distribution (Coulson et al., 1998; Friederici et al., 1993; Gunter, Stowe, & Mulder, 1997; Hagoort & Brown, 2000; Kaan 2002; Kutas & Hillyard, 1983; Münte, Matzke, & Johannes, 1997; Osterhout & Mobley, 1995), although some studies report that verb inflection errors elicit only the P600 response (Gunter & Friederici, 1999; Hagoort et al 1993; Lau, Stroud, Plesch, & Phillips, 2006; Osterhout & Nicol, 1999). Studies of subject-verb agreement processing have typically focused on person and number features, since these are more commonly encoded on the verbs of well-studied European languages. Studies of gender agreement processing have involved relations between nouns and determiners or adjectives. Gender agreement violations have been found to elicit familiar P600 effects (Hagoort & Brown, 1999) or biphasic LAN-P600 patterns

(Gunter, Friederici, & Schriefers, 2000; Barber & Carreiras, 2005). In sum, current evidence suggests that agreement violations involving person, number, and gender features elicit broadly similar ERP responses, with variation in the responses elicited by different features restricted to quantitative differences in the timing and amplitude of the P600 (Barber & Carreiras, 2005). However, to our knowledge no studies have investigated whether the ERP responses to agreement violations display the same gradient effects that have been observed in the phonetic and semantic domains, as a function of the degree of similarity between the expected and the observed word form.

The question of whether violations of one or more agreement features lead to gradient ERP effects is relevant to the theoretical questions that we opened with. If different agreement features are checked by separate sub-processes, each of which yields an ERP response when it encounters a violation, then we should expect the responses to simultaneous violations of different agreement features to be additive. Furthermore, if the checking of agreement relations involves a top-down process that compares the incoming form with an set of expected features, similar to mechanisms that have been proposed in the semantic and phonetic domains, then we expect to find smaller ERP responses to agreement violations in which the incoming form shares more features with the expected form.

There is independent motivation for a notion of degree of overlap or similarity in agreement features. In discussions of paradigm structure and syncretism in morphological theory it has been argued that closer cells in a morphological paradigm, i.e., those that share a greater number of morphological features, are more likely to behave alike with respect to allomorphy and stem selection (e.g., Anderson, 1992; Halle & Marantz, 1993; Stump, 2001; Williams, 1994; Wunderlich, 1996). For example, in richly inflected languages it often occurs that two cells in an inflectional paradigm are filled by the same form, and in such cases it is commonly the case that the two cells are neighbors in the paradigm. Thus, the general pattern that emerges is that inflectional forms that share more features in common are more likely to behave alike and to resemble each other than those that do not share common features. However, such generalizations leave open the question of whether the distance between a pair of forms in a morphological paradigm maps onto graded measures of incongruence in a way that predicts differences in ERP responses.

1.3 Subject-verb Agreement in Hindi

The current study investigates ERP responses to correct and incorrect subject-verb agreement configurations in Hindi in order to assess whether ERPs elicited by agreement violations vary as a function of the nature and/or the number of incorrect agreement features. In addition, the behavioral effects of varying the type and number of mismatching agreement features are examined by using on-line and off-line judgment tasks. Hindi future tense morphology is well suited to exploration of these questions, since it preserves a complete paradigm in which all future tense verb forms are inflected for person, number, and gender on a single word. This is rare among well-studied languages, which typically separate person and number onto an auxiliary, and number and gender onto a participle (indeed, Hindi does this in non-future tenses). This property of Hindi makes it possible to place all three independent feature-value mismatches within the same word. In the ERP experiment we varied the number and nature of feature mismatches across experimental conditions, which included two 1-feature violations (Gender, Number), and two 2-feature violations (Gender/Number, Person/Gender). In subsequent behavioral tests we added a simple Person feature violation. Combined Person/Number violations could not be presented in this paradigm due to a syncretism between 1st person plural and 3rd person plural agreement.

Hindi future tense morphology follows a regular inflectional paradigm. Two sample future tense verb forms are shown in Table 1, which presents the forms in *Devanagari* script, its phonetic form, a morpheme-by-morpheme gloss, and a translation. Future tense inflection consists of two syllables. Person is marked exclusively on the first syllable, Number is marked on both syllables, and Gender is marked on the last syllable. A more detailed explanation of the relevant morphological forms, as well as their phonetic and orthographic representations, can be found in Appendix A.

गायेगा			गायेंगे		
	gaayegaa			**gaayengee**	
gaay-	**e-**	**gaa**	**gaay-**	**en-**	**gee**
sing	*3rd/Singular*	*Masc/Singular*	*sing*	*3rd/Plural*	*Masc/Plural*
	"He will sing"			"They (masc.) will sing"	

Table 1: Sample Hindi verb agreement forms in the future tense, illustrating the words in the Devanagari script used in the current studies, in roman script, plus morphological parses and glosses.

In summary, the current study makes use of the flexible Hindi verb agreement paradigm in order to investigate whether morphosyntactic processing shows the same gradient effects of overlap between encountered and expected forms that have been observed in other domains, including auditory processing and higher-level semantic processing. ERP is an appropriate technique, since it does not require any task beyond reading for comprehension, and allows direct inferences about the timing of processes with millisecond-level resolution. The results can in turn shed light on questions about the mechanisms that underlie agreement processing.

2. Experiment 1: Materials and Methods

2.1 Participants

Twenty-three members of the University of Maryland community participated in this ERP study. Data from four participants were excluded due to unacceptably high levels of artifacts in the EEG recordings, and data from two further participants were excluded because of low judgment accuracy on one of the 5 conditions during the ERP study. The remaining 17 participants (6 females; mean age 23.9 years) were all healthy, native speakers of standard Hindi with no history of neurological disorder, and all were strongly right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971). Participants were primarily natives of Uttar Pradesh and Madhya Pradesh in north central India, regions where standard Hindi is the dominant language and the full verb agreement paradigm is used. In order to screen for mastery of standard Hindi agreement morphology and fluency in reading the Hindi *Devanagari* script, all participants took part in an off-line pre-test, consisting of 15 questions that addressed possible variation in grammatical forms. A number of speakers of non-standard dialects were excluded based on errors in this pre-test, and a small number of additional participants were excluded because they lacked the reading fluency needed to comprehend Hindi sentences presented in an RSVP paradigm. All participants whose data are included in the analyses passed all screening tests. All participants gave informed consent and were paid $10/hour for their participation, which lasted 2.5-3 hours, including set-up time.

2.2 Design and Materials

The experiment manipulated the congruency of subject verb agreement across 5 conditions, taking advantage of the fact that Hindi future tense verbs are marked separately for person, gender, and number. The critical verb was always the final word of a sentence-initial adverbial clause, thereby reducing the risk of ERP effects associated with sentence-final wrap-up. A grammatical control condition with 3rd person masculine singular verb agreement was contrasted with four ungrammatical conditions, consisting of two conditions that mismatched in just one agreement feature (Gender, Number) and two conditions that mismatched in two agreement features (Gender/Number, Person/Gender). We tested only 4 of the 7 possible combinations of ungrammatical agreement features due to constraints on the length of the experiment, and because combined Person/Number violations are not possible, due to the above-mentioned syncretism between 1st person plural and 3rd person plural. All conditions were identical except for the agreement suffixes on the critical verb.

The critical verb always appeared as the sixth word of the sentence and either matched or mismatched with a nominative subject noun phrase (NP) that appeared in word positions 2-4. The first word of the subject NP was a demonstrative determiner, which is both natural and common in Hindi, due to the lack of definite determiners. The demonstrative was followed by an adjective-noun sequence that was distinctively marked as masculine singular on at least one word. Nominative masculine nouns and adjectives in Hindi follow one of two patterns, either with number specific variants, e.g., *rasoiyaa / rasoiyee,* 'cook' sg/pl., *dublaa/dublee,* 'thin' sg./pl., or with number invariant forms, e.g., *jaj,* 'judge', sg. or pl., *samajhdaar*, 'sensible' sg. or pl. In all experimental items at least one of the adjective or the noun was distinctively marked as singular. Hindi has canonical verb-final word order, and consequently the critical verb was placed inside a sentence-initial adverbial clause, such that it would not appear in sentence-final position.

150 sets of 5 conditions were distributed across 5 lists in a Latin Square design. Three sentence-initial subordinators each appeared in 50 sets of items, *haalanki*, 'although', *chunki*, 'since', i.e., 'due to the fact that …', *jab*, 'when', i.e., 'at the time that …'. Thus, participants read 30 target sentences with correct agreement and 120 target sentences with incorrect agreement. The filler items included 120 sentences in a past-tense frame, 60 of which displayed incorrect tense on the embedded verb, and 180 additional filler items using the same subordinators as above. 45 of the filler items displayed various noun-phrase internal agreement errors. Thus, each sentence list contained a total of 225 correct and 225 incorrect sentences. Despite the fact that all target sentences contained a masculine 3rd person singular subject NP and 80% of target sentences contained agreement violations, it is unlikely that participants could anticipate the agreement violations, due to the 2:1 filler:target ratio and the fact that the filler sentences included no verb agreement violations and often had a 3rd person masculine singular subject NP. A sample item is shown in Table 2, along with the five verb forms corresponding to each of the five conditions, in both Roman script and the Devanagari script used in the experiment. Each condition is differentiated from the others by vowel diacritics found above the symbols' line, except for the Person/Gender violation, which is the only one of

the five conditions that marks the vowel change by changing the penultimate character in addition to vowel diacritics. A more detailed description of how agreement morphology is conveyed in Devanagari script can be found in Appendix A.

<table>
<tr><td colspan="5">Haalanki vo safiraa sangitkaar gaanaa gaayegaa lekin Shrotaa hoslaa nahii baRhaayenge
although that crazy.masc musician song sing.fut.3rd.sng.masc but listeners morale NEG enhance
“Although that crazy musician will sing a song, the listeners won’t boost his morale.”</td></tr>
<tr><td>गायेगा
gaayegaa
Correct
Agreement</td><td>गायेगी
gaayegii
Gender
Violation</td><td>गायेंगे
gaayengee
Number
Violation</td><td>गायेंगी
gaayengii
Gender + Number
Violation</td><td>गाऊंगी
gaayuungii
Person + Number
Violation</td></tr>
</table>

Table 2: Sample set of experimental items for Experiment 1. A Latin Square design employed all 5 variants of the sentence above, differing only in the inflected verb shown in boldface for each condition.

2.3 Procedure

Participants were comfortably seated in a dimly lit testing room around 100 cm in front of a computer monitor. Sentences were presented one word at a time in black letters on a white screen in the 30 pt Devanagari font Shusha. Each sentence was preceded by a fixation cross. Participants pressed a button to initiate presentation of the sentence, which began 1000 ms later. Each word appeared on the screen for 400 ms, followed by 200 ms of blank screen. Pre-testing showed that the 600 ms SOA was a comfortable reading rate for native speakers of Hindi, and that faster presentation rates led to substantial difficulty for many participants. The last word of each sentence was marked with a period, and 1000 ms later a question mark prompt appeared on the screen. Participants were instructed to read the sentences carefully without blinking and to indicate with a button press whether the sentence was an acceptable Hindi sentence. Feedback was provided for incorrect responses. Each experimental session was preceded by a 12 trial practice session that included both grammatical and ungrammatical sentences. Participants received feedback and were able to ask clarification questions about the task at this time.

The experimental session itself was divided into six blocks of 75 sentences each. Breaks were permitted after each block, as necessary.

2.4 EEG recording

EEG was recorded from 30 Ag/AgCl electrodes, mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, O1/2. Recordings were referenced online to the linked average of the left and right mastoids. Additional electrodes were placed on the left outer canthus, and above and below the left eye to monitor eye movements. EEG and EOG recordings were amplified and sampled at 1 kHz using an analog bandpass filter of 0.1-70 Hz. Impedances were kept below 5 kΩ.

2.5 EEG analysis

All analyses are based upon grand averages of 1100 ms intervals surrounding the critical verb, consisting of a 100 ms pre-stimulus baseline interval and a 1000 ms post-stimulus interval. Trials with ocular and other large artifacts were rejected based on visual screening. Among the 17 included participants, the total rejection rate was 16.4% ranging from 15.0%-18.4% across conditions. A 10 Hz low-pass filter was applied to the grand average ERPs for display purposes, however all statistics were performed on unfiltered data. ANOVAs were calculated based on mean voltages within a series of 200 ms time intervals that allowed continuous tracking of the evolution of any ERP responses elicited by the target word (0-200 ms, 200-400 ms, 400-600 ms, 600-800 ms, 800-1000 ms), plus one additional time interval that was included based on reports of a left anterior negativity (LAN) in previous literature (300-500 ms).

For statistical analyses, six regions of interest (ROIs) were defined, consisting of three electrodes at each ROI: left anterior (FT7, F3, FC3), midline anterior (FZ, FCZ, CZ), right anterior (F4, FC4, FT8), left posterior (TP7, CP3, P3), midline posterior (CPZ, PZ, OZ), and right posterior (CP4, P4, TP8). Two sets of ANOVAs were performed hierarchically. A first set of ANOVAs used the within-subjects factors *condition* (5 levels), *anterior-posterior*, and *laterality* (left/midline/right), with follow-up analyses based on comparisons between pairs of conditions. At the 600-800 ms time interval that marked the peak of the P600 a second set of ANOVAs was conducted that removed the person/gender condition and replaced the *condition* factor with a 2 × 2 factorial design using the factors *number* (correct/incorrect) and *gender* (correct/incorrect). This analysis allowed for an additional test of the additivity of responses to violations of different agreement features. All *p*-values reported below reflect the application of the Greenhouse-Geisser correction where appropriate to control for violations of the sphericity assumption (Greenhouse & Geisser, 1959), together with the original degrees of freedom. Due to the large number of possible interactions, we report as significant only those interactions for which follow-up analyses yielded significant contrasts within the levels of the interacting factors.

3. Results

3.1 Acceptability Judgment Task

Overall accuracy on the acceptability judgment task (n=17) was 93.6%, with individual condition means as follows: grammatical control 90%, Gender violation 93%, Number violation 91%, Gender/Number violation 98%, Person/Gender violation 99%.

3.2 ERPs

Figure 1 shows topographic scalp maps of the mean difference between each violation condition and the control condition at successive 200 ms intervals. The grand averaged waveforms for all five conditions at 9 selected electrodes distributed across the scalp are shown in Figure 2. Visual inspection indicates that all four agreement violation conditions elicited a broadly distributed posterior positivity with a peak amplitude in the

600-800ms interval, and that this positivity had an earlier onset and larger amplitude in the Person/Gender violation condition than in the three other violation conditions. The P600 effect continued to the 800-1000 ms interval, albeit with a reduced amplitude. Visual inspection provided no evidence of the left anterior negativity (LAN) response that has been observed in a number of ERP studies of morphosyntactic violations, and this was confirmed by statistical analyses at the 300-500ms interval that is characteristic of this response. Therefore, in what follows we report ANOVA results only for the 200 ms intervals shown in Figure 1.

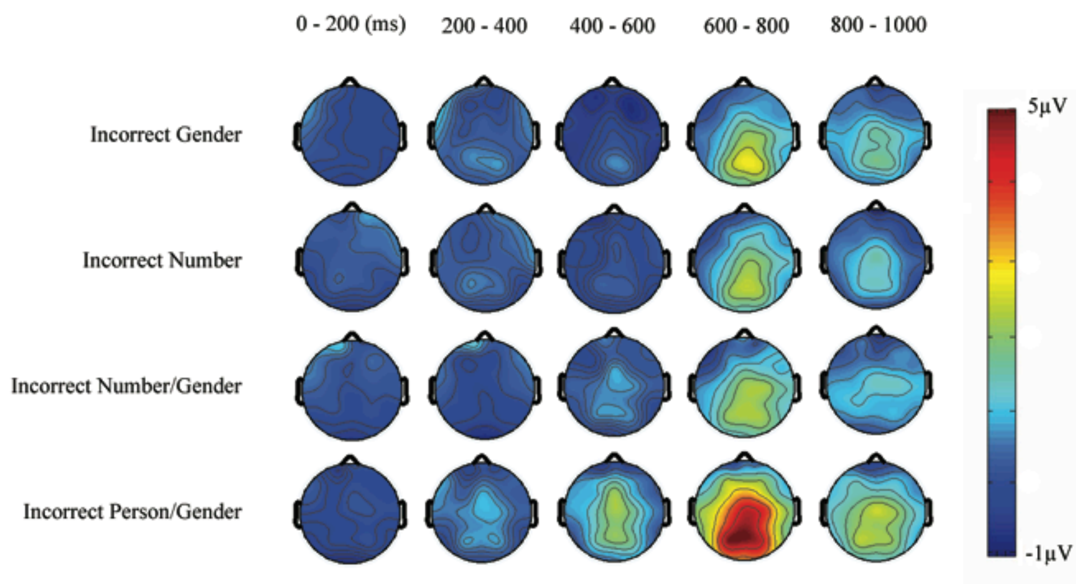

Figure 1: Topographic scalp voltage maps, showing the grand average difference between each grammatically incorrect condition and the control condition at each successive interval following the critical verb.

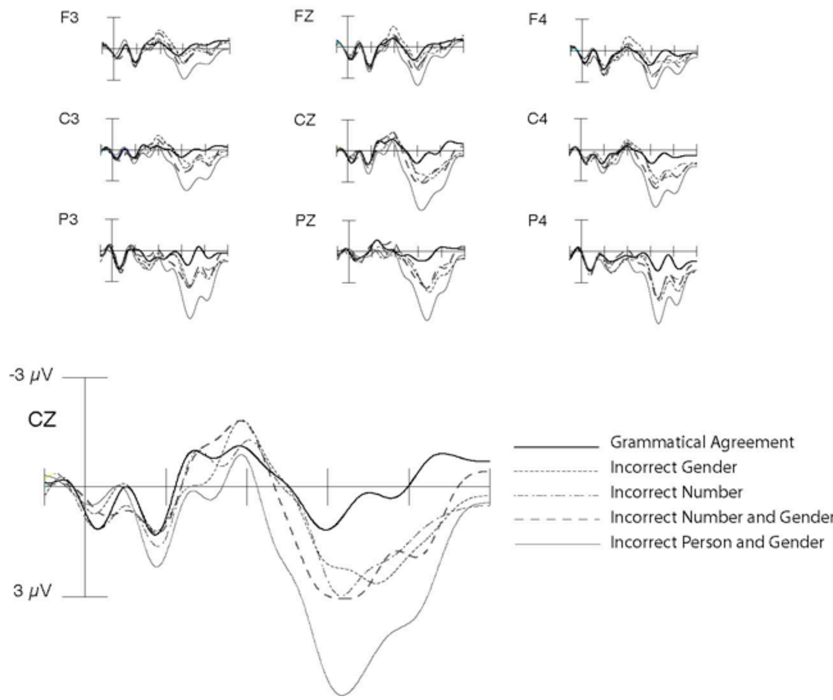

Figure 2: ERP waveforms showing grand average responses elicited by the critical verb at 9 selected electrode in all 5 conditions. Central electrode CZ is shown in greater detail.

In the comparison of the responses to the critical verb across all five conditions no reliable effects involving the condition factor were observed at the 0-200 ms interval or the 200-400ms interval.

At the 400-600 ms interval the overall ANOVA showed no main effect of condition, but showed a three-way interaction between the factors condition, anteriority, and laterality, $F(8,128) = 2.93$, $p < .01$, reflecting the fact that the P600 response to agreement violations started in this interval and had a central posterior focus. Subsequent pairwise comparisons among conditions showed a significant main effect of condition in the comparison of the Person/Gender violation with each of the four other conditions (Person/Gender vs. Control, $F(1,16) = 5.37$, $p < .05$; Person/Gender vs. Number/Gender, $F(1,16) = 3.89$, $p < .07$; Person/Gender vs. Number, $F(1,16) = 4.34$, $p < .06$; Person/Gender vs. Gender, $F(1,16) = 3.96$, $p < .07$. There were no other reliable differences between conditions at this time interval.

At the 600-800 ms interval the overall ANOVA showed a main effect of condition, $F(4,64) = 10.35$, $p < .0001$, as well as interactions of condition and anteriority,

$F(4,64) = 8.54$, $p < .0001$, and of condition and laterality, $F(8,128) = 2.78$, $p < .01$. These interactions reflect the posterior central focus of the P600 effect. Due to the large number of comparisons involved in this analysis, the *F*-values and significance levels for all main effects and interactions are presented in Table 3. The table shows that the main effect of condition and the condition × anteriority interaction and the condition × laterality interaction were significant for the comparison of each individual violation condition with the control condition, and also for the comparison of the Person/Gender condition with all other violation conditions. No other comparisons showed reliable differences in this analysis. Visual inspection of the topographic plots in Figure 1 suggests that the response to the grammatical violations has a similar central posterior scalp distribution across conditions, and the interactions with the topographic factors *anteriority* and *laterality* shown in Table 3 are consistent with this conclusion. As a further test of the similar scalp distribution of the P600 across conditions Table 4 shows results from comparisons of the same conditions, separated into anterior and posterior regions in a two-way ANOVA with the factors condition and laterality. This analysis confirms that all mismatching conditions showed similar topographic distributions, with a posterior focus for the P600, as reflected in Table 4 in generally larger *F*-values for the effect of condition at posterior regions.

cond (1,16)
cond × *ant* (1,16)
cond × *lat* (2,32)
cond × *ant* × *lat* (2,32)

	Control	Incorrect Gender	Incorrect Number	Incorrect Number/Gender
Incorrect Gender	6.460* 11.12** 3.56* n.s.			
Incorrect Number	7.18* 7.16* 3.59* n.s.	n.s. n.s. n.s. n.s.		
Incorrect Number/Gender	12.67** 10.39** 3.30† n.s.	n.s. n.s. n.s. n.s.	n.s. n.s. n.s. n.s.	
Incorrect Person/Gender	32.11*** 50.33*** 7.14** n.s.	11.58** n.s. n.s. n.s.	11.30** 6.50* 2.93† n.s.	32.71*** 13.76** 5.31* n.s.

Table 3: Results of pairwise comparisons of ERPs elicited by all conditions at 600-800 ms after the critical verb, all regions combined. Each cell in the table reflects the

F-values for the main effects and interactions in the comparison of two conditions. A legend showing the effects and degrees of freedom presented in each cell appears at the top right corner of the table. † .05 < p < .1; * p < .05; ** p < .01; *** p < .001.

Anterior / Posterior	Control	Incorrect Gender	Incorrect Number	Incorrect Number/Gender	Incorrect Person/Gender
Control		n.s. 2.60†	3.75† 3.51†	5.44* 3.88*	18.60*** 6.85**
Incorrect Gender	10.39** 3.82*		n.s. n.s.	n.s. n.s.	7.40* n.s.
Incorrect Number	10.20** 2.52†	n.s. n.s.	*cond* (1,16) *cond* × *lat* (2,32)	n.s. n.s.	5.33* 2.91†
Incorrect Number/Gender	19.54*** n.s.	n.s. n.s.	n.s. n.s.		17.92*** 4.33*
Incorrect Person/Gender	45.81*** 4.67*	15.00** n.s.	19.47*** n.s.	47.47*** 4.4.40*	

Table 4: Results of pairwise comparisons of ERPs elicited by all conditions at 600-800 ms after the critical verb, separating anterior and posterior regions. Each cell in the table reflects the F-values for the main effects and interactions in the comparison of two conditions. Comparisons at anterior electrodes are shown in the upper right half of the table, and comparisons at posterior regions in the lower left half. A legend showing the effects and degrees of freedom presented in each cell appears at the top right corner of the table. † .05 < p < .1; * p < .05; ** p < .01; *** p < .001.

At the 800-1000ms interval the overall ANOVA showed a marginally significant main effect of condition, $F(4,64) = 2.28$, $p < .07$. Pairwise comparisons revealed significant or marginally significant effects of condition in the comparison of Person/Gender vs. Control, $F(1,16) = 6.14$, $p < .05$, Person/Gender vs. Number/Gender, $F(1,16) = 3.43$, $p < .09$, and Gender vs. Control, $F(1,16) = 3.76$, $p < .08$. As in the previous time intervals, the effects were strongest in the posterior midline region.

Additionally, a 2 × 2 ANOVA was performed on four of the five conditions during the 600-800ms time window, treating ±gender agreement and ±number agreement as fully crossed factors. A main effect of number agreement was revealed ($F(1,16) = 7.02$,

$p < .05$), together with a marginally significant main effect of gender agreement ($F(1,16) = 4.27$, $p < .06$). In an analysis that focused on posterior electrode sites, where the P600 response is normally stronger, we observed a main effect of number agreement ($F(1,16) = 8.93$, $p < .01$) and of gender agreement ($F(1,16) = 9.25$, $p < .01$), plus a marginally significant interaction of gender and number ($F(1,16) = 4.08$, $p < .07$). In a corresponding analysis at anterior electrode sites a main effect of number agreement was found ($F(1,16) = 4.87$, $p < .05$).

3.3 Discussion

As expected, all of the Hindi subject-verb agreement violations in this study elicited a P600 response with that latency and central-posterior focus that are characteristic of this component. The particular interest of the current study was in whether the P600 component would show systematic variation as a function of the number of type of incorrect agreement features, in particular whether simultaneous violations of multiple agreement features would elicit a larger P600 response than violations of a single agreement feature. Larger P600 responses to violations of multiple agreement features are expected if agreement processing is governed by top-down mechanisms that prime the features of the expected verb form. Furthermore, additive effects of combined agreement violations are expected if agreement processing involves a series of independent agreement checking processes.

The ERP results did not show that 2-feature violations elicited a consistently larger P600 than did 1-feature violations. In particular, the response to the combined Gender/Number violation did not differ from the responses to individual Gender and Number violations. However, the combined Person/Gender violation did elicit a stronger P600 response than all other violations, and therefore it is important to clarify the cause of this difference before drawing theoretical conclusions.

There are a number of possible reasons why the combined Person/Gender violation elicited a larger P600. First, it may reflect the fact that this was a violation of two agreement features rather than one. This would imply that combined violations sometimes yield larger or additive responses, and sometimes do not, as in the case of the combined Gender/Number violation. Second, it may reflect the fact that violations of the Person feature are more salient to Hindi speakers than violations of other agreement features. Violations of the Person feature may be orthographically more salient to Hindi speakers than violations of other features, as can be seen in Table 2, although this is not representative of all verbs used in the study, due to the syllabic nature of Devanagari script (see Appendix A for details). Alternatively, violations of the Person feature may be more salient due to the greater cognitive or linguistic prominence that Person has been claimed to enjoy (Carminati, 2005; Greenberg, 1963). Third, it is possible that the violation involving the Person feature was stronger due to the relative rarity of Person violations in the experiment. The Person feature was incorrect in one experimental condition (6.6% of all sentences), the Number feature was incorrect in two experimental conditions (13.3% of sentences), and the Gender feature was incorrect in three experimental conditions (20% of sentences). Some previous studies indicate that infrequent violations elicit larger P600 responses (Coulson et al., 1998; Hahne & Friederici, 1999), although these findings have resulted from more dramatic frequency manipulations than those used here.

The results of the ERP study make it difficult to decide among these different accounts of the stronger response to the Person/Gender violation. We therefore conducted a pair of additional behavioral studies that were designed to determine whether the Person/Gender violation has a special status because it is a 2-feature violation, because of the salience of the Person feature, or because of the rarity of Person violations in the ERP study. We return to further discussion of the theoretical implications of the ERP results in the General Discussion.

4. Experiment 2: Acceptability Judgment Tasks

4.1 Participants

Twelve native speakers of standard Hindi (5 females, mean age of 25;4) from the University of Maryland community participated in this study, none of whom had participated in Experiment 1. Participants were selected based on dialect background in the same manner as in Experiment 1, gave informed consent and were paid $10 for their participation, which lasted around 1 hour.

4.2 Design and Materials

This study consisted of two acceptability judgment tasks that differed in the mode of sentence presentation. In one task participants judged sentences that were presented on a computer screen in an RSVP procedure. This task was designed to be as similar as possible to the presentation mode used in the ERP study, and aimed to gather reaction time data on the relative speed with which different agreement violations are detected. A second task involved an off-line pencil-and-paper rating procedure that aimed to assess whether speakers judge some agreement violations to be more severe than others.

Both tasks included six conditions, consisting of the 5 conditions used in Experiment 1 plus a new condition in which only the Person feature mismatched the subject NP. Materials for both tasks were monoclausal sentences derived from the sentence-initial adverbial clauses that contained the violations in the Experiment 1 materials. Consequently, the critical verb always appeared in sentence-final position. In the grammatically correct condition the critical verb showed correct 3rd person singular masculine agreement. This was contrasted with three conditions in which the verb mismatched in one feature (Person, Gender, and Number violations), and two conditions in which the verb mismatched in two features (Person/Gender, Number/Gender). Consequently, the Person and Number features were each violated in two conditions, and the Gender feature in three conditions. The two other logically possible combinations of mismatching features could not be tested, due to the syncretism of 1st and 3rd person plural forms in Hindi. All conditions were identical except for the agreement suffixes on the critical verb.

In the on-line judgment task, 60 sets of 6 conditions were distributed across 6 lists in a Latin Square design, such that each participant read 10 sentences with correct agreement and 50 sentences with incorrect agreement. The filler items included 120 sentences in a past-tense frame, 30 of which displayed incorrect tense on the embedded verb. Thus, each list contained a total of 80 correct and 100 incorrect sentences. An example sentence, with its Devanagari script, can be seen in Table 5. A more detailed description of how agreement morphology is conveyed in Devanagari script can be found in Appendix A.

Vo mashahur gaayak shanivaar shaam ko jalase me bahut saare geet **gayegaa.**
that famous.masc singer Saturday evening on function in many many songs ***sing.fut.3rd.sng.masc***
"That famous singer will sing many, many songs at a function on Saturday evening."

गायेगा	गायेगी	गायेंगे	गाऊंगा	गायेंगी	गाऊंगी
gaayegaa	**gaayegii**	**gaayengee**	**gaayuungaa**	**gaayengii**	**gaayuungii**
Correct Agreement	Gender Violation	Number Violation	Person Violation	Gender + Number Violation	Person + Gender Violation

Table 5: Sample set of experimental items for Experiment 2. A Latin Square design employed all 6 variants of the sentence above, differing only in the inflected verb shown in boldface for each condition.

In the off-line judgment task 30 sets of 6 conditions were distributed across 6 lists in a Latin Square design, such that participants judged 5 sentences with correct agreement and 25 with incorrect agreement. These items were also based on sentences used in Experiment 1, but were not the same items used in the on-line judgment task. The filler items included 30 sentences in a past-tense frame, 10 of which displayed incorrect tense on the embedded verb. Thus, each sentence list contained a total of 25 correct and 35 incorrect sentences.

4.3 Procedure

In the on-line judgment task sentences were presented following the same procedure used in Experiment 1. Participants were asked to give a yes/no acceptability judgment immediately after reading the sentence-final verb. Feedback was not provided for incorrect responses. Each experimental session was preceded by a brief practice session that included both grammatical and ungrammatical sentences. The off-line judgment task was administered after completion of the on-line task. Participants were instructed to rate the acceptability of each sentence on a paper questionnaire by circling the appropriate rating on a one to five scale, with one being the worst, and five the best.

4.4 Results and Discussion

The average ratings, accuracy, and response times in the off-line and on-line tasks are shown in Table 5.

Condition	Off-line Task (n=12)		On-line Task (n=10)		
	Rating	SD	Accuracy	Response Time	SE
Control	4.76	0.77	88%	1542	90
Gender Error	2.46	1.5	85%	1676	89
Number Error	2.75	1.7	76%	1792	82
Person Error	1.44	0.96	95%	1188	84
Person/Number Error	1.15	0.36	96%	1194	81
Gender/Number Error	2.06	1.3	86%	1582	96

Table 5: Experiment 2, mean ratings in the off-line acceptability rating task and mean response times in the on-line judgment task. Acceptability ratings are on a five-point scale from 1 (lowest) to 5 (highest).

The results show convergence between the off-line ratings and the accuracy and response times in the on-line task. On all measures the two conditions involving Person feature violations were judged differently from the other agreement violation conditions.

The Person violation conditions received lower acceptability ratings in the off-line task, and were judged more accurately and more rapidly in the on-line judgment task.

In the off-line rating task, paired sample t-tests showed that the control condition was rated higher than all other conditions, $p < .0001$ for the comparison of the control condition with each of the other five conditions ($t(11) = 10.106$ for Gender, $t(11) = 5.096$ for Number, $t(11) = 14.274$ for Person, $t(11) = 22.619$ for Person/Gender, and $t(11) = 9.002$ for Number/Gender). Furthermore, the Person error condition and the Person/Number error condition were rated significantly lower than the three other ungrammatical conditions: Person vs. Gender, $t(11) = 3.86$, $p < .01$; Person vs. Number, $t(11) = 3.39$, $p < .01$; Person vs. Gender/Number, $t(11) = 2.82$, $p < .05$; Person/Number vs. Gender, $t(11) = 5.70$, $p < .001$; Person/Number vs. Number, $t(11) = 5.06$, $p < .001$; Person/Number vs. Gender/Number, $t(11) = 3.49$, $p < .01$. No other differences in ratings were significant.

The results for the on-line rating task are based on 10 participants. 2 participants who gave accurate judgments but made very slow responses (typically 4s or longer) were excluded from the analysis, and all other trials with response times greater than 3500 ms were excluded. Pairwise *t*-tests for paired samples revealed that the response times for the Person and Person/Gender violations did not differ, but that these conditions were significantly or marginally significantly different from all other conditions: Person vs. Control, $t(9) = 2.08$, $p < .07$; Person vs. Gender, $t(9) = 3.12$, $p < .05$; Person vs. Number, $t(9) = 2.91$, $p < .05$; Person vs. Number/Gender, $t(9) = 1.95$, $p < .09$; Person/Gender vs. Control, $t(9) = 2.17$, $p < .06$; Person/Gender vs. Gender, $t(9) = 2.91$, $p < .05$; Person/Gender vs. Number, $t(9) = 3.40$, $p < .01$; Person/Gender vs. Number/Gender, $t(9) = 2.25$, $p < .06$.

In sum, the results of the two behavioral tasks reinforce and clarify the findings from the ERP study. First, violations of the Person feature stand out as particularly salient for Hindi speakers, who judge these violations more accurately and more quickly, and rate them as less acceptable than other agreement violations. This was true both for violations of the Person feature alone and for combined violations of Person and Gender. Second, the special status of the Person feature violations cannot be due to an experiment-specific effect of the frequency of Person violations within the experimental materials, a possibility that could not be excluded in Experiment 1. Violations of the Person feature and the Gender feature occurred with equal frequency in the judgment tasks, and yet judgments clearly differed. Third, we again found no additive effect of 1-feature vs. 2-feature violations. Ratings and response latencies for the combined Number/Gender violation did not reliably differ from those for individual violations of Number and Gender. Similarly, judgments of the combined Person/Gender violation did not reliably differ from those for the Person violation alone.

5. General Discussion

Taken together, the results of Experiments 1 and 2 suggest that Hindi speakers process agreement violations involving two incorrect features in the same way that they process violations involving one incorrect feature, thus providing evidence for a lack of additivity in the responses to violations of multiple features. This can be seen clearly in the ERP results for the Gender and Number features in Experiment 1, where the P600 elicited by a combined violation did not differ from the P600 elicited by the violations of the two features individually. However, the interpretation of this finding was undermined by uncertainty over the cause of the larger P600 elicited by the Person/Gender violation. The results of Experiment 2 suggest that the stronger response to the Person/Gender violation is caused by the special status of Person feature violations more generally, and not by the fact that the Person/Gender violation involved two incorrect features. In what follows we discuss the theoretical implications of the non-additivity in the P600 response to multiple violations, the possible causes of the special status of the Person feature, and possible reasons for the lack of a LAN effect in response to the agreement violations in this study.

5.1 Lack of a Feature Distance Effect

We interpret the results from the ERP study to show that the amplitude of the P600 response to agreement violations is not directly affected by the number of incorrect agreement features, at least for the combinations of Hindi agreement features that were tested in our studies. In other words, the distance between the correct (expected) and incorrect (observed) verb form, measured in terms of morphological features, does not directly affect P600 amplitude.

This lack of a feature distance effect contrasts with evidence from other domains of language processing, where stimuli that are equally improbable or incongruous have been found to elicit differential ERP responses as a function of their similarity to the expected stimulus. In the domain of semantic processing, Federmeier & Kutas found that implausible sentence-final nouns elicited greater or smaller N400 responses as a function of their semantic overlap with a highly expected target word, independent of the plausibility or cloze probability of the incongruous words (Federmeier & Kutas, 1999). Similar effects have been found in the domain of acoustic and phonetic processing, where the amplitude of the mismatch negativity response has been found to increase when the standard and deviant sounds differ in multiple dimensions rather than just one dimension (Paavilainen et al., 2001).

In discussing the feature distance effect in semantic processing, Federmeier and Kutas (1999, p. 387) suggest that the semantic features of a highly expected word are 'pre-activated' based on contextual cues, and that this has the consequence that words that share features with the pre-activated word are partially primed, and hence elicit smaller N400 responses. Therefore, the absence of a feature distance effect for subject-verb agreement in the current study might indicate a lack of prediction or a lack of priming for morphosyntactic features. If the agreement-relevant features of the subject NP are not primed in advance of the presentation of the verb, then this could explain why the P600 is unaffected by the number of incorrect agreement features.

Even if it is true that the Hindi speakers in our study did not engage top-down predictive mechanisms in the processing of subject-verb agreement, there is a need for caution in generalizing beyond the current findings to conclusions about morphosyntactic processing in general, and it is also important to reconcile our findings with previous results that have been presented as evidence for top-down mechanisms in agreement processing. First, our study focused on clauses with 3rd person singular masculine subject NPs, a choice that was necessary in order to allow for maximally distinctive agreement violations. 3rd person masculine singular may have a special status as a 'default' agreement form, such that this feature combination does not engage top-down processes whereas other combinations do. This would be compatible with findings by Lukatela and colleagues, who reported that in sequential lexical decision tasks adjectives facilitate the processing of morphologically congruent nouns only when the adjectives bear a non-default (i.e., non-nominative) case form (Gurjanov et al., 1985; Lukatela et al., 1982). Our conclusions are also compatible with recent reading-time results that show that the processing of agreement-neutral auxiliaries like *would* and *must* is affected by probabilistic number biases (Solomon & Pearlmutter, submitted). Although Solomon and Pearlmutter interpret their finding as evidence for a 'forward tracking' model of agreement processing that has a top-down component, it could also be understood to simply show that readers compute the likely inflectional features of all auxiliaries in a bottom-up fashion, regardless of whether they show distinctive agreement marking. In addition, one auditory judgment study in French shows latency variation in the detection of different agreement violations that are presented as an effect of the number of incorrect features (Lambert & Kail, 2001). However, the results are equally compatible with an explanation in terms of the type of incorrect features, and are thus compatible with our findings.

A further variable that should be considered when drawing conclusions about top-down mechanisms in sentence processing is the parser's ability to predict when an anticipated word or set of morphemes will appear in the word order of a sentence. Subject NPs and the verbs that agree with them may be adjacent in the word order of a sentence, or they may be separated by intervening words. In our Hindi sentences the verb always appeared in clause-final position, and was separated from the subject NP by at least one intervening word. Also, Hindi shows variation across tenses in the realization of agreement, as is true in many other languages. We tested Hindi future tense verbs, which realize person, gender, and number agreement on a single synthetic form, but Hindi present tense forms consist of a participle that marks number and gender agreement and an auxiliary that marks person and number agreement. Although our target sentences all contained future tense verbs, it was not possible for speakers to reliably predict the tense of the verb until they reached the verb itself. For this reason it will be important in future work to test whether top-down mechanisms in morphosyntactic processing are modulated not only by the predictability of inflectional features, but also by the parser's ability to anticipate exactly when the features will appear in the input.

A further possibility that should be considered is that the P600 may have shown no feature distance effect in our study because the P600 reflects agreement checking processes that are fundamentally different from the semantic integration processes that are reflected in the N400 response. If the P600 elicited by agreement violations reflects a process that simply evaluates grammatical well-formedness, then we should not expect the P600 to distinguish between verb forms that show one vs. multiple incorrect agreement features. This contrasts with a widespread view of the N400 as reflecting a process that uses information derived from incoming words to update the semantic representation of a sentence. This is a process that can clearly be facilitated in cases where features of the target word are primed, such that those features can be semantically integrated before the target word is presented, thereby reducing the N400 elicited by the target word.

A second primary goal of this research was to investigate whether agreement checking is carried out by different individual sub-processes for different types of agreement features. Some previous behavioral and electrophysiological studies have found processing differences between individual agreement features, particularly number and gender (Barber & Carreiras, 2005; de Vincenzi, 1999). Indeed, our ERP study found weak evidence in support of a timing difference between number and gender processing that was previously reported by Barber and Carreiras. In the 800-1000 ms time interval the P600 effect remained marginally significant in the Person violation condition, but this was not the case in the Number violation condition. This suggests a possible difference between number and gender processing, but the effect should be treated with caution, since the effect was not strong. However, if different agreement features were checked individually, then we might expect additive ERP effects in conditions involving simultaneous violations of multiple features. This prediction was not borne out, and thus we find no evidence for independent processing of different agreement features. This conclusion is consistent with the results from a sequential lexical decision study of agreement processing in Serbo-Croatian that found no effect of individual vs. multiple feature violations (Lukatela et al., 1987).

5.2 Why Person Is Stronger

A clear result of Experiment 1 is that the P600 elicited by a combined Person/Gender violation is significantly larger than the P600 response elicited by other agreement violations. The results from the judgment tasks in Experiment 2 indicate that Person violations and combined Person/Gender violations are processed similarly, and therefore suggest that the larger P600 response in Experiment 1 was due to the special status of Person feature violations, rather than to any enhanced response to violations of multiple agreement features. We suggest that the differences between Person violations and violations of Number and Gender agreement reflect the increased salience of Person violations, and there are a number of Hindi-specific and cross-linguistic findings that corroborate the special status of the Person feature.

A first possibility is that Person violations are more salient for orthographic reasons, due to the details of Devanagari script, as shown in Appendix A, Tables 6-7. Within the paradigm of certain verbs the person form is orthographically the most marked, although this is not clearly true for all verbs, due to the varying representation of the syllable that contains the Person agreement morpheme in Devanagari script. A second possibility is that the larger P600 response reflects a greater degree of cognitive salience accorded to the Person feature, i.e., the person feature violation may have a special status at the level of semantic representation.

A third possibility is that the Person feature has a privileged status at the level of morphosyntax, as suggested by converging evidence from language processing and language typology. Based on results from a series of self-paced reading studies Carminati argues that Italian speakers assign special importance to person agreement information in resolving the reference of an ambiguous null subject (Carminati, 2005). Carminati appeals to an earlier typological generalization by Greenberg that also gives privileged status to Person agreement (Greenberg, 1963). Based on detailed analyses of agreement morphology and pronoun systems in typologically diverse languages, Greenberg argued that languages only encode number and gender agreement if they also encode person agreement. Relatedly, it has been found that even very simple pronominal systems reserve unique forms for person while doing away with number and gender distinctions (e.g., Pirahã; Thomason and Everett, 2001). Thus, independent evidence from a number of different sources points to the importance of Person as a privileged agreement feature, which may explain the especially large P600 elicited by the Person/Gender violation in Experiment 1, and the distinctive results for violations involving the Person feature in Experiment 2.

5.3 Lack of LAN Effect

Although the Hindi agreement violations elicited a consistent P600 response, we did not find evidence for the left anterior negativity (LAN) response that has been observed in response to morphosyntactic violations in some other studies, typically with a latency of 300-500 ms. A survey of previous ERP studies shows variability in whether a subject-verb agreement violation elicits a LAN-P600 combination (Coulson et al., 1998; Hagoort & Brown, 2000; Osterhout & Mobley, 1995; Vos et al., 2001) or just a P600 (Gunter &

Friederici, 1999; Hagoort et al., 1993; Lau et al., 2006; Osterhout & Nicol, 1999). In some studies, the presence of the LAN has been found to vary as function of auditory vs. visual presentation mode (Hagoort & Brown, 2000) or as a function of working memory capacity (Vos, Gunter, Kolk, & Mulder, 2001). Therefore, it is difficult to conclude that the LAN response is a direct reflection of detection of morphosyntactic errors (Friederici, 2002) or of 'failure to bind' (Hagoort, 2003), since subject-verb agreement violations are reliably detected across studies but these violations do not reliably elicit a LAN. To our knowledge, no existing model of ERPs in sentence processing explains the variability that has been observed across studies in the presence of a LAN response to morphosyntactic violations that are superficially similar. It is possible that the presence of a LAN depends on details of the specific morphological forms tested, or on the linear relation between the subject NP and the mis-agreeing verb, or on individual differences in agreement checking mechanisms. We may speculate that the presence of the LAN may be modulated by predictive, top-down mechanisms at the morphological level in addition to the word category level (Lau et al., 2006). However, much further work will be required in order to identify the factors that govern the variability in the presence of the LAN, and the current study is not well suited to resolve this issue.

6. Conclusion

The objective of the current study was to determine whether the parser is sensitive to the degree of similarity between expected and unexpected agreement marking on a verb, with the aim of better understanding the contribution of top-down and bottom-up proceses in agreement checking. The current study capitalized on the rich verb agreement morphology of Hindi to test whether feature distance affects the size of the P600 elicited by an agreement violation. Previous studies on the processing of semantic and acoustic anomalies have demonstrated a feature distance effect of this kind, and this effect has been used as a diagnostic of a top-down priming mechanism. In varying the degree to which an anomalous verb form diverged from the expected form, we found that the P600 did not vary as a function of the number of incorrect agreement features, contrasting with findings in other domains of language processing. This finding may reflect qualitative differences between the processing of syntactic and semantic anomalies, or it may reflect differences in the use of top-down predictive mechanisms in semantic and morphosyntactic processing. The non-additive effect of combining agreement violations also shows that different agreement features are not checked independently. The results from the ERP and behavioral experiments additionally suggested that a larger P600 response is associated with violations of the Person feature than violations of other agreement features. This result that is consistent with a variety of previous linguistic and psycholinguistic results that have shown that Person has a privileged status among agreement features.

Appendix A: Hindi Verb Agreement

This appendix describes how Hindi future tense verb agreement maps onto *Devanagari* script. Although future tense verb agreement follows a singular regular paradigm, certain aspects of the agreement paradigm are realized in multple ways, due to the fact that Devanagari is a syllabic script. This is particularly relevant to the finding in our study that agreement violations involving the person feature elicited a larger P600.

Hindi future tense verb agreement is phonologically regular. Future tense verb forms consist of a verb stem and two suffixes. The first suffix carries information about person and number, and the second suffix carries information about gender and number. These suffixes apply to all verb stems, and the suffixes used in our ERP study are shown in Table 6.

Person/Number		Gender/Number	
-uun-	1st Person, Singular	*-gaa*	Masculine, Singular
-ee-	3rd Person, Singular	*-gee*	Masculine, Plural
-een-	3rd Person, Plural	*-gii*	Feminine, Sing/Plur

Table 6: The phonological forms of the agreement morphemes used in the experiment

Devanagari script is a syllabary, in which each character corresponds to a syllable. Consequently, the realization of each suffix depends on the syllable that the suffix is a part of. Sample forms for three different verb stems used in our ERP study are shown in Table 7. The gender/number suffix includes a syllable onset, and therefore each suffix is consistently realized with the same Devanagari character. In contrast, the Person/Number suffix does not include a consonant as a syllable onset, and therefore this suffix forms a syllable together with the final consonant of the verb stem (or with *-y-* if the stem ends in a vowel). Consequently, each Person/Number suffix may be realized by a different Devanagari character, according to the consonant that the suffix combines with to form a syllable. This has a potential impact upon the orthographic salience of different agreement violations. As shown in Table 6, 3rd person singular and 3rd person plural suffixes have the same vowel quality and differ only in the presence of nasalization, which is marked by a diacritic mark above the character (see the dot in the 1st person singular and 3rd person plural forms in Table 6). In contrast, 1st person singular and 3rd person singular forms contain different vowels (*-uun-* vs. *-ee-*, respectively), and therefore are realized by a different base Devanagari character, according to the syllable onset provided by the final consonant of the stem. There is a good deal of variation in the orthographic salience of the 1st person vs. 3rd person singular contrast, depending on the verb stem. For example, in Table 7 the difference between the suffixes *-uun-* and *-ee-* maps to a large orthographic contrast when they are attached to the stem *gaa* ‘sing’ and marked by the Devanagari characters for *yuun* and *yee*. On the other hand, when the same suffixes appear in the syllables *nuun* and *nee* (*maan* ‘obey’) or *tuun* and *tee* (*jiit* ‘win’) the orthographic contrast is less salient. Thus, agreement violations involving the person feature are more salient than other violations when applied to some verb stems, but not when applied to other verb stems.

-uun- (1st, Singular)	*-ee-* (3rd, Singular)	*-een-* (3rd, Plural)
मानूंगी (maanuungii)	मानेगी (maaneegii)	मानेंगी (maaneengii)
गाऊंगा (gaayuungaa)	गायेगा (gaayeegaa)	गायेंगे (gaayeengee)
जीतूंगा (jiituungaa)	जीतेगा(jiiteegaa)	जीतेंगे (jiiteengee)

Table 7: The orthographic forms for three verb stems: *jiit* 'to win', *gaa* 'to sing', and *maan* 'to obey'. The first row of the table contains feminine forms, which end in *–gii*, and the other two rows contain masculine forms, which end in *–gaa* and *–gee*.

References

Altmann, G.T.M., Kamide, Y., 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition 73, 247-264.

Anderson, S., 1992. A-morphous morphology. Cambridge University Press, Cambridge, UK.

Aoshima, S., Yoshida, M., Phillips, C., submitted. Incremental processing of coreference and binding in Japanese.

Bader, M., Lasser, I., 1994. German verb-final clauses and sentence processing: Evidence for immediate attachment. In: Perspectives on sentence processing. C. Clifton, L. Frazier, K. Rayner, eds. Erlbaum, Hillsdale, NJ, pp. 225-242.

Barber, H., Carreiras, M., 2005. Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. Journal of Cognitive Neuroscience 17, 135-153.

Bock, J.K., Cutting, J.C., 1992. Regulating mental energy: Performance units in language production. Journal of Memory and Language 31, 99-127.

Bock, J.K., Miller, C.A., 1991. Broken agreement. Cognitive Psychology 23, 45-93.

Carminati, M.N., 2005. Processing reflexes of the Feature Hierarchy (Person > Number > Gender) and implications for linguistic theory. Lingua 115, 259-285.

Chen, E., Gibson, E., F. Wolf, 2005. Online syntactic storage costs in sentence comprehension. Journal of Memory and Language 52, 144-169.

Corbett, G., 2006. Agreement. Cambridge University Press, Cambridge, UK.

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

M. Crocker, 1996. Computational psycholinguistics: an interdisciplinary approach to the study of language. Kluwer, Dordrecht, The Netherlands.

de Vincenzi, M., 1999. Differences between the morphology of gender and number: Evidence from establishing coreference. Journal of Psycholinguistic Research 28, 537-553.
Eberhard, K.M., Cutting, J.C., Bock, J.K., 2005. Making syntax of sense: number agreement in sentence production. Psychological Review 112, 531-559.
Elman, J., 1991. Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning 7, 195-226.
Federmeier, K.D., Kutas, M., 1999. Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. Cognitive Brain Research 8, 373-392.
Forchheimer, P., 1953. The Category of Person in Language. Walter de Gruyter, Berlin,.
Franck, J., Vigliocco, G., Nicol, J., 2002. Subject-verb agreement errors in French and English: The role of syntactic hierarchy. Language and Cognitive Processes 17, 371-404.
Frazier, L., Fodor, J.D., 1978. The sausage machine: a new two-stage parsing model. Cognition 6, 291-326.
Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing, Trends in Cognitive Sciences 6, 78-84.
Friederici, A.D., Pfeifer, E., Hahne, A., 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations. Cognitive Brain Research 1, 183-192.
Gibson, E., 1998. Linguistic complexity: locality of syntactic dependencies. Cognition 68, 1-76.
Greenhouse, S., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95-112.
Gunter, T., Friederici, A.D., 1999. Concerning the automaticity of syntactic processing. Psychophysiology 36, 126-137.
Gunter, T., 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.
Gunter, T., Stowe, L.A., Mulder, G., 1997. When syntax meets semantics. Psychophysiology 34, 660-676.
Gurjanov, M., Lukatela, G., Moskovljevic, J., Savic, M., Turvey, M.T., 1985. Grammatical priming of inflected nouns by inflected adjectives. Cognition 19, 55-71.
Hagoort, P., 2003. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. Neuroimage 20, S18-S29.
Hagoort, P., Brown, C.M., 1999. Gender electrified: ERP evidence on the syntactic nature of gender processing. Journal of Psycholinguistic Research 28, 715-728.
Hagoort, P., Brown, C.M., 2000. ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sequences and rapid serial visual presentation. Neuropsychologia 38, 1531-1549.
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, 439-484.
Hagoort, P., Wassenaar, M., Brown, C., 2003. Syntax-related ERP-effects in Dutch. Cognitive Brain Research 16, 38-50.

Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. Journal of Cognitive Neuroscience 11, 194-205.
Hale, J., 2003. The information conveyed by words in sentences. Journal of Psycholinguistic Research 32, 101-123.
Halle, M., Marantz, A., 1993. Distributed morphology and the pieces of inflection. In: The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger, K.L. Hale, S.J. Keyser, eds. MIT Press, Cambridge, MA, pp. 111-176.
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.
Kamide, Y., Altmann, G.T.M., Haywood, S.L., 2003. The time-course of prediction in incremental sentence comprehension: Evidence from anticipatory eye-movements. Journal of Memory and Language 49, 133-156.
Kimball, J., 1975. Predictive analysis and over-the-top parsing. In: Syntax and semantics, vol. 4, J. Kimball, ed. Academic Press, New York, pp. 155-179.
Kutas, M., Federmeier, K.D., 2004. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences 4, 463-470.
Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: Brain potentials reflect semantic anomaly. Science 207, 203-205.
Kutas, M., Hillyard, S.A., 1983. Event-related brain potentials to grammatical errors and semantic anomalies. Memory & Cognition 11, 539-550.
Lambert, L., Kail, M., 2001. Le traitement en temps reel des marques morphologiques d'accord dans des phrases en Français. L'Année Psychologique 101, 561-592.
Lau, E.F., Stroud, C., Plesch, S., Phillips, C., 2006. The role of structural prediction in rapid syntactic analysis. Brain and Language 98, 74-88.
Levänen, S., Hari, R., McEvoy, L., Sams, M., 1993. Responses of the Human Auditory Cortex to Changes in one versus two Stimulus Features. Experimental Brain Research 97, 177-183.
Levy, R., submitted. Expectation-based syntactic comprehension.
Lukatela, G., Kostic, A., Todorovic, D., Carello, C., Turvey, M.T., 1987. Type and number of violations and the grammatical congruency effect. Psychological Research 49, 37-43.
Lukatela, G., Moraca, J., Stojnov, D., Savic, M., Katz, L., Turvey, M.T., 1982. Grammatical priming effects between pronouns and inflected verb forms. Psychological Research (Psychologische Forschung) 44, 297-311.
Münte, T.F., Matzke, M., Johannes, S., 1997. Brain activity associated with syntactic incongruencies in words and pseudo-words. Journal of Cognitive Neuroscience 9, 318-329.
Münte, T.F., Szenkuti, A., Wieringa, B.M, Matzke, M., Johannes, S., 1997. Human brain potentials to reading syntactic errors in sentences of different complexity. Neuroscience Letters 235, 105–108.
Näätänen, R., 1992. Attention and brain function. Erlbaum, Hillsdale, NJ.
Näätänen, R., Gaillard, A.W.K., Mäntysalo, S., 1978. Early selective attention effect on evoked potential reinterpreted. Acta Psychologica 42, 313–329.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I., 2001. 'Primitive intelligence' in the auditory cortex. Trends in Neurosciences 24, 283-288.

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

Nicol, J.L., Forster, K.I., Veres, C., 1997. Subject-verb agreement processes in comprehension. Journal of Memory and Language 36, 569-587.

Oldfield, J.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9, 97–113.

Osterhout, L., Holcomb, P.J., Swinney, D.A., 1994. Brain potentials elicited by garden path sentences: evidence of the application of verb information during parsing. Journal of Experimental Psychology: Learning, Memory, & Cognition 20. 768-803.

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

Osterhout, L., Nicol, J., 1999. On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. Language and Cognitive Processes 14, 283-317.

Paavilainen, P., Valppu, S., Näätänen, R., 2001. The additivity of the auditory feature analysis as indexed by the mismatch negativity: 1+1 = approximately 2, but 1+1+1 < 3. Neuroscience Letters 301, 179-182.

Pearlmutter, N.J., Garnsey, S.M., Bock, J.K., 1999. Agreement processes in comprehension. Journal of Memory and Language 41, 427-456.

Solomon, E.S., Pearlmutter, N.J., submitted. Forward versus backward agreement processing in comprehension.

Staub, A., Clifton Jr., C., 2006. Syntactic prediction in language comprehension: evidence from *either … or*. Journal of Experimental Psychology: Learning, Memory, & Cognition 32, 425-236.

Stump, G., 2001. Inflectional morphology: A theory of paradigm structure. Cambridge University Press, Cambridge, UK.

Thomason, S., Everett, D., 2001. Pronoun Borrowing. Proceedings of the Berkeley Linguistics Society 27.

Van Petten, C., Kutas., M., 1990. Interactions between sentence context and word frequency in event-related brain potentials. Memory and Cognition 18, 380-393.

Van Petten, C., Kutas, M., 1991. Influences of semantic and syntactic context on open and closed class words, Memory and Cognition 19, 95-112.

Vigliocco, G., Nicol, J., 1998. Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? Cognition 68, 13-29.

Vos, S., Gunter, T.C., Kolk, H., Mulder, G., 2001. Working memory constraints on syntactic processing: an electrophysiological investigation. Psychophysiology 30, 41-63.

Wechsler, S., Zlatic, L., 2003. The many faces of agreement. CSLI Publications, Stanford, CA.

Williams, E., 1994. Remarks on lexical knowledge. Lingua 92, 7-34.

Wunderlich, D., 1996. Minimalist morphology: the role of paradigms. In: Yearbook of Morphology 1995, G. Booij, J. van Marle, eds. Kluwer, Dordrecht, The Netherlands, pp. 93-114.