



## Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension

Ina Bornkessel-Schlesewsky<sup>a,\*</sup>, Franziska Kretzschmar<sup>a,b</sup>, Sarah Tune<sup>a</sup>, Luming Wang<sup>b</sup>, Safiye Genç<sup>c</sup>, Markus Philipp<sup>a,b</sup>, Dietmar Roehm<sup>d</sup>, Matthias Schlewsky<sup>b</sup>

<sup>a</sup> Department of Germanic Linguistics, University of Marburg, Marburg, Germany

<sup>b</sup> Department of English and Linguistics, Johannes Gutenberg-University Mainz, Mainz, Germany

<sup>c</sup> Department of German Language and Literature, Marmara University, Istanbul, Turkey

<sup>d</sup> Department of Linguistics, University of Salzburg, Salzburg, Austria

### ARTICLE INFO

#### Article history:

Available online 22 October 2010

#### Keywords:

Language comprehension  
Electrophysiology  
N400, P600  
P300  
Semantic reversal anomalies  
Semantics  
Syntax  
Verb-argument linking  
Word order  
Categorisation

### ABSTRACT

This paper demonstrates systematic cross-linguistic differences in the electrophysiological correlates of conflicts between form and meaning (“semantic reversal anomalies”). These engender P600 effects in English and Dutch (e.g. Kolk et al., 2003; Kuperberg et al., 2003), but a biphasic N400 – late positivity pattern in German (Schlewsesky and Bornkessel-Schlewsesky, 2009), and monophasic N400 effects in Turkish (Experiment 1) and Mandarin Chinese (Experiment 2). Experiment 3 revealed that, in Icelandic, semantic reversal anomalies show the English pattern with verbs requiring a position-based identification of argument roles, but the German pattern with verbs requiring a case-based identification of argument roles. The overall pattern of results reveals two separate dimensions of cross-linguistic variation: (i) the presence vs. absence of an N400, which we attribute to cross-linguistic differences with regard to the sequence-dependence of the form-to-meaning mapping and (ii) the presence vs. absence of a late positivity, which we interpret as an instance of a categorisation-related late P300, and which is observable when the language under consideration allows for a binary well-formedness categorisation of reversal anomalies. We conclude that, rather than reflecting linguistic domains such as syntax and semantics, the late positivity vs. N400 distinction is better understood in terms of the strategies that serve to optimise the form-to-meaning mapping in a given language.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

From our own day-to-day experience, we are all familiar with the diversity of the world’s languages: A typical English-speaking tourist in China or Turkey, for example, will be painfully aware of how different these languages seem to his/her native tongue. But what are the consequences of this diversity for the neurobiology of language? Is the neural language architecture essentially the same across languages or are there important language-specific differences?

With neuroscientific methods for the examination of human cognition now commonplace in many laboratories around the world, research questions such as these have started to become more and more frequent in recent years. For example, a number of studies have examined whether less-studied languages such as Chinese show similar electrophysiological or neuroanatomical pro-

cessing signatures for particular types of linguistic violations to languages such as English (e.g. Luke, Liu, Wai, Wan, & Tan, 2002; Ye, Luo, Friederici, & Zhou, 2006; Ye, Zhan, & Zhou, 2007). However, when data of this type reveal cross-linguistic differences, these are not trivial to interpret. For example, if a syntactic violation in a language like Chinese engenders an electrophysiological effect that has been assumed to reflect semantic processing in European languages, should this be interpreted as evidence for the importance of semantics in Chinese grammar? Alternatively, a result such as this could be taken to indicate that our functional interpretation of the effect in question is incorrect, since it does not generalise across languages of different types.

In this paper, we present evidence in favour of the second view by demonstrating systematic cross-linguistic variation in electrophysiological responses to a single linguistic phenomenon. On the basis of data from Turkish, Chinese and Icelandic, we show that languages fall into several distinct types with regard to their neurophysiological processing signatures for conflicts between form and meaning (“semantic reversal anomalies”). Specifically, we will argue that there are two distinct dimensions of cross-linguistic variation with regard to the neurophysiological correlates

\* Corresponding author. Address: Department of Germanic Linguistics, University of Marburg, Wilhelm-Roepke-Strasse 6A, 35032 Marburg, Germany. Fax: +49 (0)6421 2824558.

E-mail address: [iboke@staff.uni-marburg.de](mailto:iboke@staff.uni-marburg.de) (I. Bornkessel-Schlewsesky).

of reversal anomalies: whether verb–argument linking in a given language is sequence-dependent or not (reflected in modulations of N400 effects); and whether the grammatical properties of a language allow for a binary stimulus categorisation (well-formed vs. ill-formed; reflected in modulations of late positivity effects). In the following, we first introduce semantic reversal anomalies and present some initial evidence for cross-linguistic differences in the electrophysiological patterns with which they correlate. We then present a neurocognitive model of language comprehension that can derive these differences and that makes further cross-linguistic predictions.

### 1.1. Semantic reversal anomalies: Conflicts between form and meaning reveal a cross-linguistic puzzle

The processing of so-called semantic reversal anomalies has attracted a great deal of attention in recent years. These anomalies (see example 1) occur in syntactically well-formed but implausible sentences, the key feature of which is that they would be plausible with reversed role assignments (i.e. if the hearty meals in 1a were the object rather than the subject of the sentence and subject and object were interchanged in 1b). They thus induce a conflict between form (the role assignments prescribed by the grammar) and meaning (the most plausible combination of the elements in the sentence).

#### (1) a. Example of a semantic reversal anomaly in English (Kim & Osterhout, 2005)<sup>1</sup>

#The hearty meals were devouring . . .

#### b. Example of a semantic reversal anomaly in Dutch (from Hoeks, Stowe, & Doedens, 2004)

#De speer heeft de atleten geworpen.  
the javelin has the athletes thrown  
'The javelin has thrown the athletes.'

<sup>1</sup> Here and in the following, we use the hash symbol (#) to designate an implausible, but grammatically well-formed sentence. Sentence positions at which the critical event-related potential (ERP) effects were measured are underlined.

In English and Dutch, semantic reversal anomalies have been shown to consistently elicit late parietal positivities (P600 effects) in comparison to plausible control conditions (Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; van Herten, Chwilla, & Kolk, 2006; van Herten, Kolk, & Chwilla, 2005). This result was surprising because it appeared to contradict the traditional association between event-related potential (ERP) components and linguistic subdomains, according to which semantic conflicts engender N400 effects (for reviews, see Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008), whereas P600s correlate with aspects of syntactic processing (for reviews, see Gouvea, Phillips, Kazanina, & Poeppel, 2010; Hagoort, 2003; Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004). The finding of "semantic P600" effects for reversal anomalies such as (1) thus raises two questions: (a) Why do reversal anomalies not engender an N400?; and (b) Why do they engender a P600?

The finding of monophasic P600s for semantic reversal anomalies has often been interpreted as evidence for a (hitherto unexpected) dominance of semantic information in the online composition of sentence-level meaning (for discussion, see Kim & Osterhout, 2005; Kuperberg et al., 2007; van Herten et al., 2006). The logic behind this conclusion is as follows: if syntactic composition preceded semantic interpretation, sentences such as (1) should lead to a semantic mismatch at the position of the verb,

which should be reflected in an N400. The absence of an N400 suggests that no such mismatch occurred, thereby indicating that the processing system first combined the sentence constituents in the most plausible way. The mismatch between the representation generated in this way and the role assignments required by the syntactic structure is then thought to elicit the P600, though the precise functional interpretation of this effect has been debated (for a review, see Bornkessel-Schlesewsky & Schlesewsky, 2008). Thus, the absence of an N400 is typically accounted for via the assumed primacy of semantic composition, while the presence of a P600 is assumed to reflect some aspect of detecting the erroneous analysis (e.g. Kim & Osterhout, 2005; Kolk et al., 2003).

However, in spite of the reliability with which monophasic P600 effects have been reported for semantic reversal anomalies in English and Dutch, recent findings from German call into question the cross-linguistic generalisability of these results. In contrast to the results described above, Schlesewsky and Bornkessel-Schlesewsky (2009) observed N400 effects for sentences such as (2).

#### (2) Examples from Schlesewsky and Bornkessel-Schlesewsky (2009)

##### a. Irresolvable semantic reversal anomaly in German

. . . dass der Schalter den Techniker bedient  
. . . that [the switch]:NOM [the technician]:ACC  
operates  
#'. . . that the switch operates the technician.'

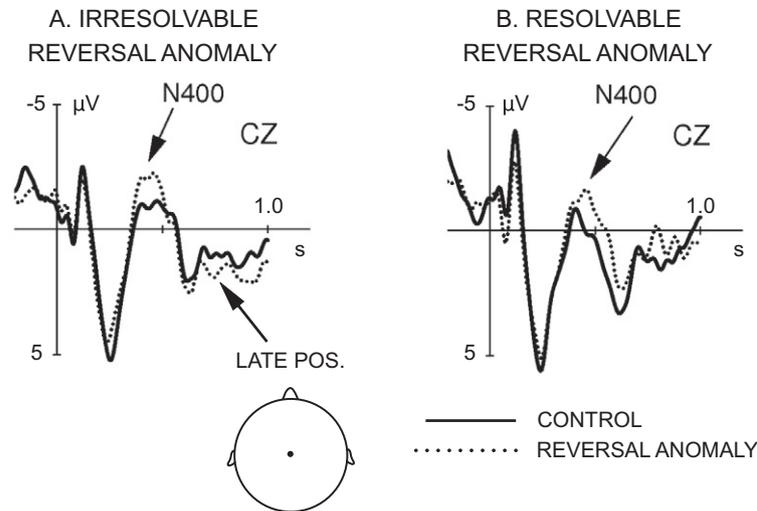
##### b. Resolvable semantic reversal anomaly in German

. . . dass Schalter Techniker bedienen  
. . . that switches technicians operate  
' . . . that technicians operate switches.'

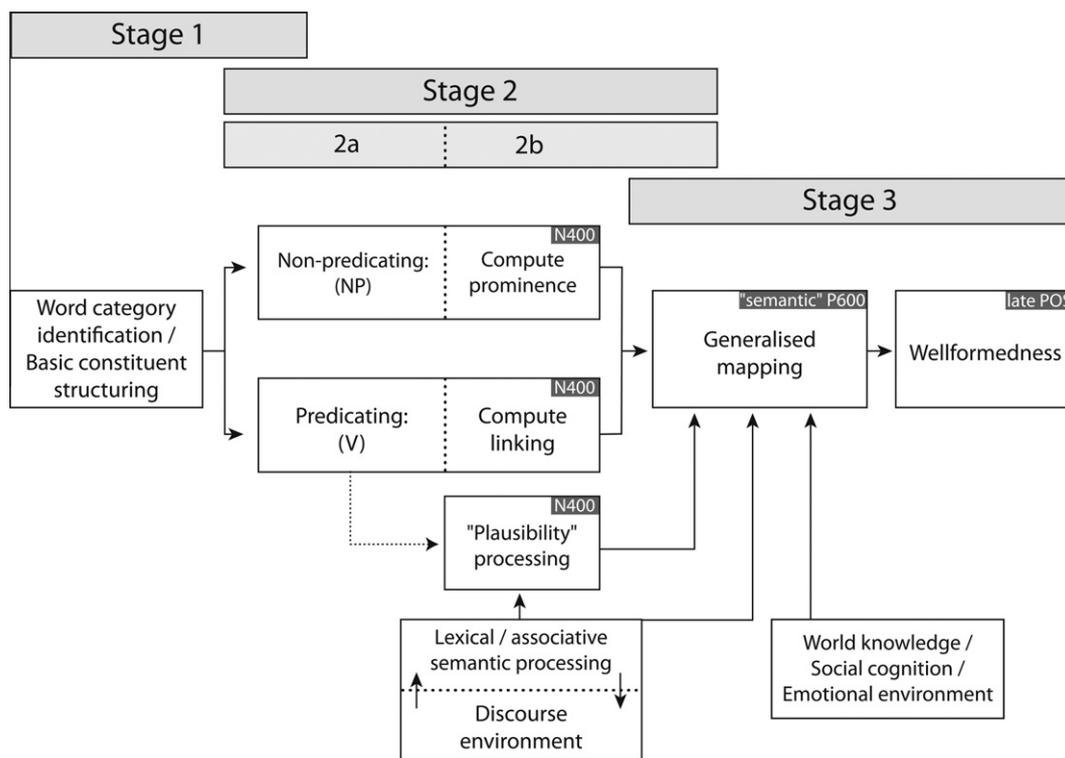
Example (2a) corresponds very closely to example (1b) in that the clause-final verb induces a semantic reversal anomaly. It contrasts with the Dutch example, though, in that role assignments are rendered unambiguous via morphological case marking rather than word order. When unambiguous case marking is absent, as in (2b), the well-known subject–first preference for German (see Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008, for a recent review) leads to a similar conflict at the position of the verb. However, since the sentence can be reanalysed towards an object–initial order, the reversal anomaly is resolvable in this case. Both sentence types in (2) engendered an N400 effect in comparison to control conditions (for 2a: dass den Schalter der Techniker bedient, lit: 'that [the switch]:ACC [the technician]:NOM operates'; for 2b: dass Techniker Schalter bedienen, lit: 'that technicians switches operate'). Sentences such as (2a) additionally engendered a late positivity, which Schlesewsky and Bornkessel-Schlesewsky (2009) attribute to the irresolvability of the conflict (see also Bornkessel-Schlesewsky & Schlesewsky, 2008; Kolk et al., 2003). The two effects are illustrated in Fig. 1.

### 1.2. Deriving cross-linguistic variation in the electrophysiological signatures for semantic reversal anomalies

As shown in the previous section, semantic reversal anomalies in German contrast with their counterparts in English and Dutch in that they do not elicit a monophasic P600 effect. How should this result be interpreted? Should it be viewed as evidence for fundamental differences in the interplay between syntactic and semantic information in these three closely related languages? An alternative to this rather drastic conclusion was offered by Schlesewsky and Bornkessel-Schlesewsky (2009), who interpret their findings within a cross-linguistic neurocognitive model of language comprehension, the extended Argument Dependency



**Fig. 1.** Illustration of the N400 effect for semantic reversal anomalies in German (adapted from Schlewsky & Bornkessel-Schlesewsky, 2009). Panel A presents the data for irresolvable reversal anomalies and panel B presents the data for resolvable reversal anomalies. Negativity is plotted upwards.



**Fig. 2.** The architecture of the extended Argument Dependency Model (eADM; from Bornkessel-Schlesewsky & Schlewsky, 2009b).

Model (eADM Bornkessel & Schlewsky, 2006; Bornkessel-Schlesewsky & Schlewsky, 2008, 2009b). The relevant aspects of the model architecture are shown in Fig. 2.

In brief, the eADM posits that sentence-level interpretation involves the assignment of the two generalised semantic roles “Actor” and “Undergoer” (which are based on the Agent and Patient prototypes, respectively) to the sentential arguments. The language processing system can assign the Actor and Undergoer roles independently of the verb by referring to a cross-linguistically motivated set of syntactic and semantic information types referred to as prominence information (e.g. case marking, linear

order, animacy, definiteness/specificity, person) and their language-specific weightings (compute prominence).<sup>1</sup> Once the verb is reached, these role assignments are linked to the verb's

<sup>1</sup> Note that this conception of sentence-level interpretation resembles the assumptions of the Competition Model in several respects (for overviews of this model and the extensive body of cross-linguistic research associated with it, see Bates, Devescovi, & Wulfeck, 2001; MacWhinney & Bates, 1989). For a discussion of similarities and differences between the Competition Model and the eADM, see Bornkessel and Schlewsky (2006), Bornkessel-Schlesewsky and Schlewsky (2009b).

lexical entry which consists of a decomposed semantic representation, e.g.  $do'(x, operate'(x, y))$ <sup>2</sup> for the verb *to operate* in example 2 (compute linking). Crucially, compute linking is assumed to operate independently of lexical-semantic associations and world knowledge, which are processed in a simultaneous but separate processing step (plausibility processing). Subsequently, the outputs of linking and lexical-semantic processing are integrated with one another in a step referred to as generalised mapping. As is apparent from Fig. 2, compute prominence, compute linking, and plausibility processing typically correlate with N400 effects, whereas generalised mapping typically correlates with P600 effects. For a detailed motivation of these architectural assumptions, which are based on a wide range of cross-linguistic findings on verb-argument processing in simple sentences, see Bornkessel-Schlesewsky and Schlesewsky (2009b).

Within the eADM, semantic P600 effects in English and Dutch are assumed to derive from a problem in the generalised mapping step, which arises when compute linking and plausibility processing produce incompatible outputs. In (1b), for example, the javelin and the athlete are assigned the Actor and Undergoer roles, respectively, on the basis of their sequential position in the sentence (compute prominence/compute linking). They are also both plausible arguments of the verb *to throw*, though with reversed role assignments (plausibility processing). When this conflict is detected by generalised mapping, a P600 results. Crucially, there is no conflict in compute linking, since the only information type that is relevant for this processing step in English and Dutch is the sequential order of the arguments (i.e. in a simple active sentence, when argument A precedes argument B, A is the Actor and B is the Undergoer; for empirical confirmation, cf. MacWhinney, Bates, & Kliegl, 1984).<sup>3</sup> In examples (1a) and (1b), the processing system therefore does not encounter a problem until the generalised mapping step (for extensive discussion, see Bornkessel-Schlesewsky & Schlesewsky, 2008).

In a language such as German, by contrast, argument linking cannot rely simply upon linear position. Rather, a range of sequence-independent information types including case marking and animacy need to be taken into account (cf. Bornkessel-Schlesewsky & Schlesewsky, 2008; MacWhinney et al., 1984). Hence, a processing problem already occurs in compute linking in examples (2a) and (2b) due to the conflict between case, word order and animacy information: in (2a), nominative case calls for an actor interpretation of *the switch*, while animacy information calls for an actor interpretation of *the technician*; in (2b), the actor interpretation of *the switch* is supported by word order (which is drawn upon in the

absence of unambiguous case marking), again conflicting with the animacy-based preference to interpret the technician as the actor. Both sentences therefore engender an N400 effect. The additional late positivity in (2a) is attributed to the well-formedness assessment step (cf. Fig. 2), which recognises the irresolvability of the processing problem and the resulting decrease in sentence acceptability.

These assumptions about sequence-dependence vs. sequence-independence in verb-argument linking in English vs. German are supported by investigations conducted within the scope of the Competition Model (Bates, McNew, MacWhinney, Devescovi, & Smith, 1982). In a comparative study of English, German and Italian, MacWhinney et al. (1984) showed that, for native speakers of English, word order is the dominant cue to sentence interpretation (rank order of cues for English: word order > animacy, agreement). In German, by contrast, animacy played a much more important role in determining which participant in a simple, transitive sentence was chosen as the actor (rank order of cues for German: animacy > agreement > word order) – note that this order applies for sentences in which case marking is ambiguous; when case marking is unambiguous, it is the dominant cue (Kempe & MacWhinney, 1999). Thus, for English, word order dominates interpretation even in the face of conflicting cues: speakers choose the first noun phrase (NP) in an NP-V(erb)-NP structure as the actor in over 80% of all cases even when this interpretation is contradicted by agreement, i.e. when the second NP agrees with the verb (MacWhinney et al., 1984); animacy also has only a very minor effect, with NP<sub>INANIMATE</sub>-V-NP<sub>ANIMATE</sub> structures leading to 86% of actor-first interpretations (McDonald, 1986). Notably, these empirical preferences correlate strongly with independent measures of “cue validity” in a given language. The validity of a cue (e.g. word order) is high when that cue is both highly available (i.e. often present in the input) and highly reliable (i.e. seldom misleading). Our claim is therefore that linking can be characterised as sequence-dependent when the validity of the order cue is extremely high and other cues play only a minor role.

In summary, the eADM accounts for the processing of semantic reversal anomalies in the following way. The absence of an N400 for reversal anomalies in English and Dutch is explained by the fact that (a) there is no processing problem in the plausibility processing step in view of the semantic relatedness between the arguments and the verb; and (b) there is no processing problem in compute linking because linking is “vacuously satisfied” by linear order as a highly dominant information source. The monophasic P600 observed in these languages is attributed to a mismatch between the output of compute linking and plausibility processing which registers in the generalised mapping step. In languages such as German, by contrast, in which linking draws upon multiple, sequence-independent information sources – including animacy – semantic reversal anomalies already lead to a linking conflict and, thereby, to an N400. When this conflict is irresolvable, an additional late positivity reflects a violation of well-formedness requirements. Thus, the eADM posits that the typical “semantic P600” pattern, i.e. monophasic P600s observable in response to semantic reversal anomalies, is a language-specific pattern that is restricted to languages of the English/Dutch type. Importantly, this proposal predicts that the German pattern should also be observable in other languages with similar linking characteristics, a hypothesis that was tested in the present study.

### 1.3. The present study: a cross-linguistic investigation of semantic reversal anomalies

As described in the preceding section, the eADM predicts that languages with sequence-independent verb-argument linking characteristics should show N400 effects for semantic reversal

<sup>2</sup> This decomposed lexical representation, termed the verb's logical structure (LS), is adopted from Role and Reference Grammar (see Van Valin, 2005).

<sup>3</sup> As pointed out by an anonymous reviewer, MacWhinney et al. (1984) in fact found an effect of animacy in all of the three languages that they examined (English, German and Italian). One might therefore wonder whether the formulation that word order is the only information type that is relevant for compute linking in English is too strong. On the one hand, it is clear that animacy does play a role in English: as in other languages of the world, transitive sentences most frequently occur with an animate actor and an inanimate undergoer (Jäger, 2007). However, rather than determining which argument is interpreted as the actor and which as the undergoer (as, for example, in Chinese), animacy serves to establish how prototypical a particular argument is as either actor or undergoer (for a detailed discussion, see Bornkessel-Schlesewsky & Schlesewsky, 2009b). Indeed, MacWhinney et al. (1984) themselves note that: “Of the three languages we studied here, only English relied heavily on word order as a near-categorical cue to sentence interpretation” (p. 148). Notably, while animacy accounted for approximately 1% of the variance in actor judgments in their data for English, it accounted for a little over 20% of the data for German. We thus assume that, though animacy may have a small effect in English via prototypicality (for further empirical evidence, see Chen, West, Waters, & Caplan, 2006; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005), this does not reach the threshold that would allow it to influence the N400 as a correlate of compute linking when it comes to determining “who is acting on whom” (particularly in the NP-V-NP structures under discussion in the literature on reversal anomalies in English, in which MacWhinney and colleagues' data indeed show a virtually categorical choice of the first noun phrase as actor).

anomalies (followed by a late positivity, as an index of a well-formedness violation). It thus crucially assumes that the findings from German that were reported by [Schlesewsky and Bornkessel-Schlesewsky \(2009\)](#) are not an exception, but rather form part of a cross-linguistically stable pattern which differs from that of the English/Dutch type. In order to test this hypothesis, Experiment 1 examined Turkish, a language which shares a number of characteristics with German (e.g. morphological case marking, flexible word order, verb-final sentence structures). Experiment 2 extended this line of investigation by studying Chinese, a language with flexible word order but no morphological case marking and only a limited number of verb-final sentence constructions. A comparison of the results from Turkish, Chinese and German, therefore allowed us to determine whether the different patterns in German and English/Dutch should be attributed to differences in morphological case marking (which is present in German, but almost entirely absent in English and Dutch) or to sequence-independence vs. sequence-dependence in linking. Note that, for both Turkish and Chinese, the sequence-independence of verb-argument linking is supported by investigations within the Competition Model: [Bates et al. \(2001; Table 4\)](#) list the rank order of cues for Turkish as case > animacy > word order and for Chinese as animacy > word order. Finally, Experiment 3 used Icelandic to examine whether the two distinct electrophysiological signatures for semantic reversal anomalies could be elicited in a single language via different verb types with distinct linking characteristics.

The remainder of the paper is organised as follows. Sections 2 and 3 present Experiments 1 (Turkish) and 2 (Chinese), respectively. Subsequently, in Section 4, the data from German, Turkish and Chinese are contrasted in an interim discussion. Experiment 3 on Icelandic is then reported in Section 5, followed by a general discussion in Section 6.

## 2. Experiment 1: Turkish

### 2.1. Relevant properties of Turkish

Turkish is an Altaic language with agglutinating morphology and consistently head-final characteristics (i.e. verbs follow their arguments, postpositions follow noun phrases, head nouns follow relative clauses). While the basic word order is subject–object–verb (SOV), constituent ordering is flexible and permutations are permitted. Turkish also has morphological case marking. In all of these respects, Turkish is thus more similar to German than to English and Dutch, in spite of the fact that it belongs to a different language family.

Two other features of Turkish are relevant to the present study. These concern the distribution of accusative case marking and the phenomenon of “word order freezing”. In contrast to German, direct objects are only marked with accusative in Turkish when they are interpreted as definite/specific. This phenomenon, which is known as “differential object marking” (DOM) (see [Aissen, 2003; Bossong, 1985](#)), is illustrated in (3).

(3) Differential object marking in Turkish (adapted from [Erguvanli, 1984](#), pp. 21–22)

- a. Adam kitap okuyor.  
man book read-PROG  
'The man is reading a book. (~The man is book-reading.)'
- b. Adam kitabı okuyor.  
man book-ACC read-PROG  
'The man is reading the book.'

Crucially, the presence or absence of accusative case marking has important consequences for Turkish word order: deviations from the basic SOV order are only permitted when the object is case marked (4a), otherwise word order freezing occurs (4b).

- 
- (4) Object case marking and word order freezing in Turkish (from [Erguvanli, 1984](#), pp. 21–22)
- a. Kitabı adam okuyor.  
book-ACC man read-PROG  
'The man is reading the book.'
- b. \*Kitap adam okuyor.  
Book man read-PROG
- 

Note that, while example (4b) is designated as ungrammatical by the asterisk, this ungrammaticality only arises under the (only plausible) reading that the man is reading the book. The reverse (and implausible) interpretation that the book is reading the man, by contrast, is grammatical. Example (4b) therefore has all the characteristics of a semantic reversal anomaly.

In order to examine the hypothesis that semantic reversals in Turkish should engender an N400 rather than a P600, Experiment 1 contrasted sentences such as (4b) (reversal anomalies) and (4a) (unambiguously case marked controls). We chose to examine reversal anomalies such as (4b), which show word order freezing, rather than unambiguously case marked reversal anomalies in order to test the additional prediction that the presence or absence of a linking problem in reversal anomalies should be due to the overall linking characteristics of a particular language rather than to the properties of the construction currently being processed. In other words, we predict an N400 (linking) – late positivity (well-formedness) pattern, rather than a monophasic P600 for Turkish sentences such as (4b) because linking in Turkish in principle needs to consider a variety of information types rather than simply linear order (see [Bates et al., 2001](#)).<sup>4</sup>

### 2.2. Experimental design and hypotheses

The four critical sentence condition employed in Experiment 1 are illustrated in [Table 1](#).

As is apparent from the table, the first two conditions (AI-CON and AI-REV) correspond to the examples in (4) that were described in the preceding section: the unambiguously case marked AI-CON served as the plausible control condition, whereas AI-REV instantiated the critical reversal anomaly. Furthermore, we examined

<sup>4</sup> In fact, certain word order restrictions in Turkish can be overcome by animacy, thus attesting to the fact that word order and animacy interact in this language. As shown by examples (ia) and (ib), indefinite subjects (e.g. a star) must be left-adjacent to the verb (as in ia). However, this restriction does not hold when the indefinite subject is animate (a woman in ic/id).

(i) Word order, definiteness and animacy in Turkish (examples from [Demiral, 2008](#))

- a. Gökyüzünden bir yıldız kaydı.  
sky-ABL one star slid  
'A star slid through the sky'
- b. \*Bir yıldız gökyüzünden kaydı.  
one star sky-ABL slid
- c. Merdivenden bir kadın kaydı.  
stair-ABL one woman slid  
'A woman slid from the stairs'
- d. Bir kadın merdivenden kaydı.  
one woman stair-ABL slid  
'A woman slid from the stairs'

two additional conditions (IA-CON and IA-REV), in which the animacy relations were reversed (i.e. the inanimate argument was the plausible Actor rather than the animate argument). This allowed us to test whether the electrophysiological response to the reversal anomaly is influenced by the animacy hierarchy (i.e. by the fact that animate arguments are more prototypical Actors than inanimate arguments). To the best of our knowledge, all previous studies on reversal anomalies that employed an animacy manipulation induced the anomaly via atypical role assignments (as in AI-REV). Additional filler sentences (see materials section below) served to balance the number of subject- and object-initial sentences in the experiment.

As already noted in the preceding section, we expected to observe an N400 effect for the reversal anomaly AI-REV in comparison to the control condition AI-CON in Experiment 1. This hypothesis is based on the observation that verb-argument linking in Turkish resembles that in German in that it is not sequence-based; rather, a range of information types need to be taken into account. Furthermore, if the effects for reversal anomalies are tied closely to verb-specific animacy constraints rather than to the animacy hierarchy in general, the same pattern should also be observable in the comparison between conditions IA-REV and IA-CON. Finally, we predicted that any N400 effects should be followed by a late positivity in view of the ill-formedness of the reversal anomalies.

## 2.3. Materials and methods

### 2.3.1. Participants

Twenty-two native speakers of Turkish (8 women; mean age: 27.1, range 17–38 years) participated in Experiment 1 after giving written informed consent. All were healthy, right-handed and had normal or corrected-to-normal vision and good auditory acuity. Five additional participants were excluded from the final data analysis due to excessive EEG artefacts.

### 2.3.2. Materials

The experimental materials were constructed on the basis of 80 noun pairs consisting of an animate and an inanimate noun. For each pair, two verbs were selected: one which unambiguously required the animate noun to be the Actor and one which unambiguously required the inanimate noun to be the Actor (see Table 1). Thus, eighty versions of the four critical conditions shown in Table 1 were constructed (320 sentences in total). These were subdivided into two lists of 160 sentences each (40 sentences per condition, two from each lexical set). Each list was interspersed with an additional 320 filler sentences in a pseudo-randomised manner. The fillers included variations in animacy and definiteness

**Table 1**

Sentence examples and mean acceptability ratings for the four critical conditions in Experiment 1 (Turkish). Standard deviations (by participants) are given in parentheses. Abbreviations: AI – verb calls for an animate Actor and an inanimate Undergoer; IA – verb calls for an inanimate Actor and an animate Undergoer; CON – control condition; REV – reversal anomaly; ACC – accusative case.

| Condition | Example   | Acceptability (%) |
|-----------|---|-------------------|
| AI-CON    | Haber-i subay duydu.<br>news-ACC officer heard<br>'The officer heard the news.'             | 98.5 (2.9)        |
| AI-REV    | Haber subay duydu.<br>news officer heard<br>'#The news heard the officer.'                  | 2.7 (5.0)         |
| IA-CON    | Subay-ı haber sevindirdi.<br>officer-ACC news made.happy<br>'The news pleased the officer.' | 99.2 (1.9)        |
| IA-REV    | Subay haber sevindirdi.<br>officer news made.happy<br>'#The officer pleased the news.'      | 2.3 (3.8)         |

of the noun phrases and also served to ensure that an equal number of subject and object-initial sentences occurred in the study. Each participant was presented with a single list once and list assignment was counterbalanced across participants.

All sentences were recorded by a trained, native speaker of Turkish (a former newsreader) and digitally recorded with a sampling rate of 44.1 kHz and a 16-bit resolution. Sentences were checked for naturalness by a native speaker of Turkish and re-recorded when necessary.

### 2.3.3. Procedure

Participants were seated in a dimly lit, sound attenuated room at a distance of 100 cm from a 17 inch computer screen. Each trial began with the presentation of a fixation asterisk in the centre of the screen. After 500 ms, a sentence was presented via loudspeakers. The asterisk remained on the screen for another 1000 ms after sentence offset. Five hundred milliseconds later, participants performed two tasks: an acceptability judgement (maximal response time: 2000 ms) and a probe detection task, in which they judged whether a single word that was presented in the centre of the screen had occurred in the preceding sentence or not (maximal response time: 3000 ms). Participants responded to both tasks by pressing one of two buttons on a response box. The assignment of "yes" and "no" responses to the left and right buttons was counterbalanced across participants.

Participants were asked to fixate the asterisk throughout the duration of its presentation (from 500 ms before sentence onset to 1000 ms after sentence offset) and to avoid movements and eye blinks during this time. The experimental session began with a short practice run, after which participants were presented with 8 blocks of 60 sentences each. Between blocks, participants took short breaks. Including electrode preparation, experimental sessions lasted approximately 3.5 h.

### 2.3.4. EEG recording and preprocessing

The EEG was recorded from 25 Ag/AgCl electrodes, which were fixed at the participant's scalp by means of an elastic cap (Electrocap International, Eaton, OH). The EEG was amplified using a Twente Medical Systems DC amplifier (Enschede, The Netherlands) and digitised with a sampling rate of 250 Hz (ground: AFz). The electrooculogram (EOG) was monitored via electrodes placed above and below a participant's right eye (vertical EOG) and at the outer canthi of both eyes (horizontal EOG). The left mastoid served as the reference electrode, but EEG signals were referenced to linked mastoids offline. Electrode impedances were kept below 4 k $\Omega$ .

In order to exclude slow signal drifts, the EEG data were filtered with 0.3–20.0 Hz band pass offline. Automatic and manual rejections were carried out to exclude periods containing artefacts (EOG, movement or technical) from the data analysis. Trials for which the probe detection task was answered incorrectly were also excluded from further analysis. ERPs were calculated per participant, electrode, and condition from –200 to 1200 ms relative to the onset of the critical verb, before grand averages were computed over all participants.

### 2.3.5. Data analysis

For the acceptability judgement and probe detection tasks, mean error rates per condition were subjected to a repeated measures analysis of variance (ANOVA) involving the condition factors ANOMALY (REV: sentence contained a reversal anomaly vs. CON: sentence was a non-anomalous control condition) and ANIMACY (AI: verb called for an animate Actor and inanimate Undergoer vs. IA: verb called for an inanimate Actor and an animate Undergoer) and the random factors participants ( $F_1$ ) and items ( $F_2$ ). The

reaction times were not analysed as they were not directly time-locked to the critical sentence position.

For the ERPs, we computed repeated measures ANOVAs involving the factors ANOMALY and ANIMACY and the topographical factor region of interest (ROI). Separate analyses were performed for lateral and midline electrodes. Lateral ROIs were defined as follows: left-anterior: F3, F7, FC1, and FC5; left-posterior: CP1, CP5, P3, and P7; right-anterior: F4, F8, FC2, and FC6; right-posterior: CP2, CP6, P4, and P8. The midline electrodes (Fz, FCz, Cz, CPz, Pz, and POz) each formed a ROI of their own. The statistical analyses were carried out hierarchically, i.e. only interactions that were at least marginally significant ( $p < 0.07$ ) were resolved. Huynh-Feldt (H-F)-corrected significance values are reported whenever a statistical analysis included a factor with more than one degree of freedom in the numerator (Huynh & Feldt, 1970).

As this is the first ERP study in Turkish to have undertaken a manipulation of this type and it is therefore difficult to specify pre-defined time windows, the time windows for the statistical analysis were chosen on the basis of visual inspection of the data.

## 2.4. Results

### 2.4.1. Behavioural data

Accuracy rates for the acceptability judgement and probe detection tasks were very high (higher than 97% in all conditions for the judgement task and higher than 98% in all conditions for the probe detection task). Acceptability ratings for all four conditions are given in Table 1 (note that for the analysis of the accuracy rates, an “unacceptable” judgement was coded as correct for the REV-conditions).

For the acceptability judgement, a repeated measures ANOVA revealed a significant main effect of ANOMALY ( $F_1(1, 21) = 7.47$ ,  $p < 0.02$ ;  $F_2(1, 79) = 7.30$ ,  $p < 0.01$ ). Thus, in spite of the very high accuracy rates for all conditions, the control conditions were judged more accurately than the reversal anomalies (98.86% vs. 97.50%).

The analysis of the accuracy rates for the probe detection task revealed no significant main effects or interactions.

### 2.4.2. ERP data

Fig. 3 shows grand average ERPs time-locked to the critical clause-final verb for the reversal anomalies (IA-REV and AI-REV) in comparison to their respective control conditions (IA-CON and AI-CON).

As is apparent from the figure, both reversal anomalies engendered a centrally distributed negativity between approximately 350 and 800 ms post verb onset. Statistical analyses were therefore conducted in this time window.

For the lateral electrodes, a repeated measures ANOVA for the 350–800 ms time window revealed a main effect of ANOMALY ( $F(1, 21) = 24.14$ ,  $p < 0.0001$ ) and an interaction ROI  $\times$  ANIMACY  $\times$  ANOMALY ( $F(3, 63) = 5.22$ ,  $p < 0.01$ ). However, resolving the three-way interaction by ROI did not reveal a significant ANIMACY  $\times$  ANOMALY interaction in any region.

The analysis of the midline electrodes showed a main effect of ANOMALY ( $F(1, 21) = 26.25$ ,  $p < 0.0001$ ) and interactions ROI  $\times$  ANOMALY ( $F(5105) = 6.01$ ,  $p < 0.02$ ) and ROI  $\times$  ANIMACY  $\times$  ANOMALY ( $F(5105) = 9.65$ ,  $p < 0.001$ ). Resolving the interactions by ROI revealed significant effects of ANOMALY at all six midline electrodes (min:  $F(1, 21) = 17.74$ ,  $p < 0.001$  at Fz; max:  $F(1, 21) = 28.49$ ,  $p < 0.0001$  at CPz). The interaction ANIMACY  $\times$  ANOMALY reached marginal significance at Pz ( $F(1, 21) = 4.04$ ,  $p < 0.06$ ) and POz ( $F(1, 21) = 3.78$ ,  $p < 0.07$ ). In both cases, the interaction was due to a more reliable difference between IA-REV and IA-CON (Pz:  $F(1, 21) = 15.76$ ,  $p < 0.001$ ; POz:  $F(1, 21) = 13.04$ ,  $p < 0.01$ ) than between AI-REV and AI-CON (Pz:  $F(1, 21) = 7.83$ ,  $p < 0.02$ ; POz:  $F(1, 21) = 6.16$ ,  $p < 0.03$ ).

In summary, the ERP data in Experiment 1 revealed a central negativity for reversal anomalies vs. controls. This effect, which was observable between 350 and 800 ms post critical verb onset, resembles a typical auditory N400 in its distribution; its latency is also compatible with an N400 interpretation. Thus, reversal anomalies in Turkish pattern with those in German in showing an N400. Interestingly, the N400 effect was slightly more pronounced for the IA conditions than for the AI conditions at parietal midline sites, i.e. when the reversal anomaly was effected by a verb which called for atypical Actor-Undergoer assignments (inanimate Actor, animate Undergoer). Surprisingly, however, the N400 was

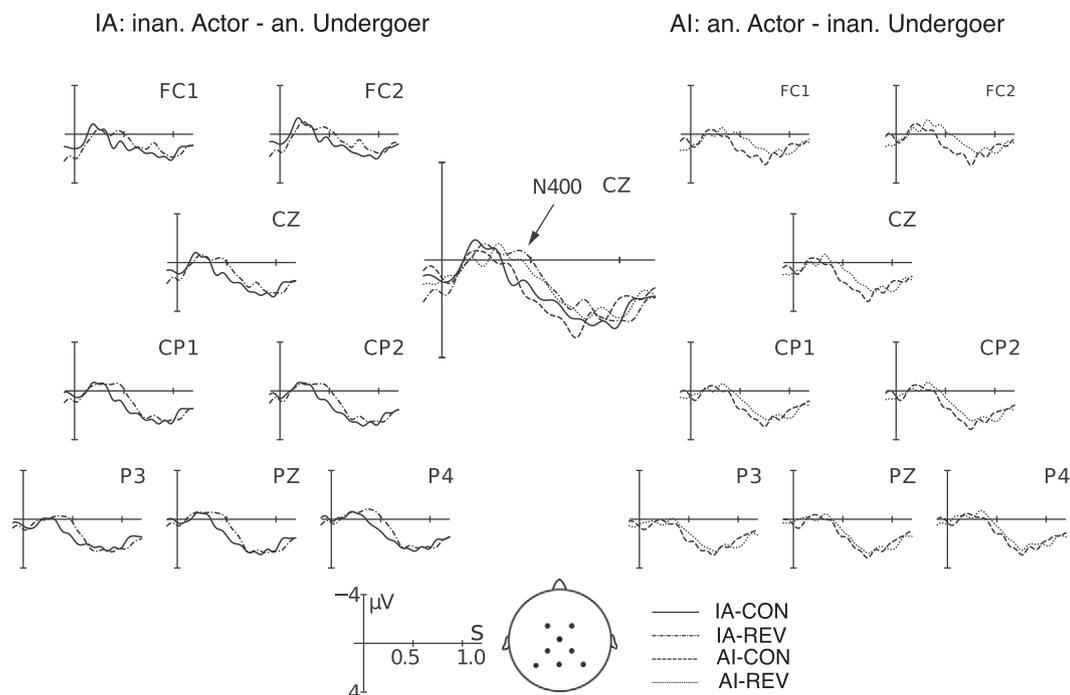


Fig. 3. Grand-average ERPs time-locked to the critical clause-final verb (onset at the vertical bar) in Experiment 1 (Turkish). Negativity is plotted upwards.

not followed by a late positivity in spite of the fact that the reversal anomaly sentences were reliably classified as unacceptable. These findings will be discussed and interpreted in more detail in the interim discussion in Section 4.

### 3. Experiment 2: Mandarin Chinese

#### 3.1. Relevant properties of Mandarin Chinese

Mandarin Chinese is a Sino-Tibetan language, which is typically described as having a basic SVO order (Sun & Givón, 1985). It is an isolating language and therefore has very little morphology (i.e. no case marking, no agreement, etc.). In these respects, it is thus rather different from both German and Turkish and perhaps somewhat more similar to English. In contrast to English, however, Chinese permits a number of word order permutations: OSV, SOV, and VOS are all possible word orders in spoken Chinese (e.g. Li & Thompson, 1981). However, since NP–NP–V orders are grammatically ambiguous between an Actor–Undergoer–Verb (SOV) and Undergoer–Actor–Verb (OSV) reading and the preference for one or the other is moreover influenced by animacy (see Li, Bates, & MacWhinney, 1993; Wang, Schlesewsky, Philipp, & Bornkessel-Schlesewsky, in press), these structures are not suited to inducing semantic reversal anomalies.

However, Chinese also has another kind of verb-final structure, in which the role assignments to the arguments are rendered unambiguous via a coverb (bǎ/bèi). Coverbs are verbs which, via a process of grammaticalisation (cf. Hopper & Traugott, 2003), came to be used in the function of prepositions. Verb-final sentences with bǎ and bèi are shown in example (5).

|     |   |
|-----|---|
| (5) | Examples of bǎ and bèi-structures (from Li & Thompson, 1981, p. 358)  |
| a.  | tā bǎ shū fàng xià lái le<br>3sg BA book put descend come CRS<br>'He/she put the books down.'               |
| b.  | wǒ bèi tā zhūi le sān tiān<br>1sg BEI 3sg chase PFV three days<br>'I was chased by him/her for three days.' |

As shown by the example in (5), bǎ and bèi both call for a sentence structure in which the verb follows both of its arguments (NP-coverb–NP-verb). They differ, however, in that bǎ unambiguously renders the first NP the Actor and the second NP the Undergoer, while these assignments are reversed with bèi (note the passive translation of 5b). These sentence structures are therefore well-suited to inducing semantic reversal anomalies in Chinese. Note that, while bǎ and bèi, may appear to serve a similar purpose to morphological case marking in a language such as German or Turkish, they are typically considered closer to prepositions than to case markers (Bisang, 1992; Li & Thompson, 1981). Sentences such as (5) can thus be considered comparable to Dutch reversal anomalies such as *De vos die op de stropers joeg...* ('The fox that at the poachers hunted...'), from Kolk et al., (2003), in which the second NP is also preceded by a preposition, and which have been shown to pattern with sentences without prepositions (e.g. example 1b) in engendering monophasic P600 effects.

#### 3.2. Experimental design and hypotheses

The four critical sentence conditions employed in Experiment 2 are illustrated in Table 2.

As is apparent from the table, the sentences show a number of similarities to those used in Experiment 1: the reversal anomaly is

induced by the clause-final verb and the two preverbal arguments differ with respect to animacy. In contrast to Turkish, however, the order of animate and inanimate NPs was always the same (though there were inanimate-initial filler sentences, see below); the assignment of Actor and Undergoer roles to the arguments was governed by the choice of coverb.

Our hypotheses were as follows. If the N400 vs. P600 distinction for reversal anomalies between German and Turkish on the one hand and English and Dutch on the other should be attributed to the presence vs. absence of a rich case morphology, Chinese should be expected to behave like English and Dutch. Under these circumstances, we should observe a P600 for the comparison between the REV and CON-conditions in Experiment 2. By contrast, if the relevant distinguishing factor is word order flexibility vs. rigidity, Chinese should pattern with German and Turkish in showing an N400 (which, according to our original predictions, should be followed by a late positivity). Note that the second prediction holds in spite of the lack of word order freedom in bǎ and bèi-structures: as shown for Turkish, the properties of a particular construction (e.g. word order freezing in the absence of morphological case marking) do not appear to determine the electrophysiological response. Rather, the effect engendered by reversal anomalies in a particular language appears to be more closely related to the overall properties of the language.

#### 3.3. Materials and methods

##### 3.3.1. Participants

Twenty-five native speakers of Mandarin Chinese (Beijing dialect) participated in Experiment 2 after giving written informed consent (13 women; mean age: 27.6 years, range 22–34). All were healthy, right-handed and had normal or corrected-to-normal vision and good auditory acuity. Three additional participants were excluded from the final data analysis due to excessive EEG artefacts.

##### 3.3.2. Materials

The experimental materials consisted of 36 sets of the four conditions shown in Table 2. Each set comprised an animate noun, an inanimate noun and two verbs: one which unambiguously rendered the animate noun the Actor and the inanimate noun the Undergoer and one which required the reversed assignments. It was further insured that all verbs were compatible with both bǎ and bèi-sentences (for details on the semantic restrictions imposed by these coverbs, see Chang, 2003; Li & Thompson, 1981). The 144 critical sentences thus resulting were interspersed with an

**Table 2**

Sentence examples and mean acceptability ratings for the four critical conditions in Experiment 2 (Mandarin Chinese). Standard deviations (by participants) are given in parentheses. Abbreviations: BEI – condition with the coverb bèi (first NP = Undergoer); BA – condition with the coverb bǎ (first NP = Actor); CON – control condition; REV – reversal anomaly.

| Condition | Example   | Acceptability (%) |
|-----------|---|-------------------|
| BEI-CON   | 侦探 被 子弹 击中了。<br>detective BEI bullet hit<br>'The detective was hit by the bullet.'    | 90.0 (10.5)       |
| BEI-REV   | 侦探 被 子弹 保存了。<br>detective BEI bullet kept<br>'#The detective was kept by the bullet.' | 7.0 (5.6)         |
| BA-CON    | 侦探 把 子弹 保存了。<br>detective BA bullet kept<br>'The detective kept the bullet.'          | 96.6 (4.8)        |
| BA-REV    | 侦探 把 子弹 击中了。<br>detective BA bullet hit<br>'#The detective hit the bullet.'           | 12.5 (8.2)        |

additional 288 filler sentences in a pseudo-randomised manner. Of the filler sentences, 144 realised a separate experimental manipulation that did not involve *bǎ* and *bèi*-structures. The remaining fillers were *bǎ* and *bèi*-sentences with an inanimate-initial argument and an animate second argument; they thereby served to counterbalance the animacy relations in the current experimental design (in which NP1 was always animate).

For purposes related to the experimental design that was realised within the filler sentences (which was concerned with aspects of information structure processing and word order in Chinese), each sentence was preceded by a context question that introduced the first NP of the experimental sentence as a discourse topic. These questions were always of the form in (6). The context questions were the same for all critical conditions. Moreover, they provide a highly natural lead into the experimental sentences as the first position in the sentence is typically described as a topic position in Chinese (Li & Thompson, 1976).

---

(6) 侦探 怎么了?  
zhēntàn zěnmē-le  
detective how-PFV  
'What is about the detective?'

---

Sentences were digitally recorded by two native speakers of Chinese (Beijing dialect), a male speaker for the context questions and a female speaker for the target sentences, using a sampling rate of 44.1 kHz and a 16-bit resolution. They were subsequently checked for naturalness by a native speaker of Chinese and re-recorded if necessary.

The 432 sentences in the experiment (144 critical sentences and 288 fillers) were presented to participants in two different randomised presentation orders.

### 3.3.3. Procedure

The experimental sessions were conducted in the same setting as Experiment 1. Each trial began with the presentation of a fixation asterisk in the centre of the screen. After 500 ms, a sentence pair consisting of a context question and a target sentence was presented via loudspeakers (inter-stimulus interval: 1000 ms). The asterisk remained on the screen for 1000 ms after the end of the target sentence. Following 500 ms of blank screen, participants were asked to judge the acceptability of the sentence (maximal reaction time: 2000 ms) and, after a further 1000 ms of blank screen, to complete a comprehension task (maximal reaction time: 4000 ms). For the comprehension task, participants decided whether a question which appeared in the centre of the screen correctly described the preceding experimental sentence or not. Comprehension questions consisted of both *bǎ* and *bèi*-sentences for all conditions, thus ensuring that participants could not simply engage in a linear matching of words from the critical sentence to the control condition. Questions requiring the answer "incorrect" included a substituted argument or verb. Participants responded to both tasks by pushing one of two buttons on a response box. The assignment of the left and right buttons to "yes" and "no" answers was counterbalanced across participants.

Participants were asked to avoid movements and eye blinks throughout the presentation of the fixation asterisk. The experimental session began with a short training session, followed by 12 experimental blocks comprising 36 sentences each, between which the participants took short breaks. The entire experimental session (including electrode preparation) lasted approximately 3.5 h.

### 3.3.4. EEG recording and preprocessing

The EEG recording and preprocessing procedures were identical to those employed for Experiment 1.

### 3.3.5. Data analysis

The data analysis was performed as for Experiment 1 with the only exception that Experiment 2 employed the factors ANOMALY (REV: sentence contained a reversal anomaly vs. CON: sentence was a non-anomalous control condition) and COVERB (BA: NP1 = Actor; BEI: NP1 = Undergoer).

## 3.4. Results

### 3.4.1. Behavioural data

The accuracy rates for the acceptability judgement and comprehension tasks were again high (>87% for all conditions in the judgement task and >93% for all conditions in the comprehension task).

For the judgement accuracy, a repeated measures ANOVA showed an interaction of ANOMALY  $\times$  COVERB ( $F_1(1, 24) = 32.46$ ,  $p < 0.0001$ ;  $F_2(1, 35) = 6.62$ ,  $p < 0.02$ ). Resolving this interaction by COVERB revealed an effect of ANOMALY for the BA-conditions ( $F_1(1, 24) = 21.36$ ,  $p < 0.001$ ;  $F_2(1, 35) = 9.08$ ,  $p < 0.01$ ) but not for the BEI-conditions ( $F_1(1, 24) = 1.17$ ,  $p > 0.28$ ;  $F_2 < 1$ ). Thus, for the BA-sentences only, judgements were less accurate for the reversal anomaly than for the control condition.

The statistical analysis of the comprehension accuracy revealed a main effect of ANOMALY which was significant by participants and marginal by items ( $F_1(1, 24) = 9.11$ ,  $p < 0.01$ ;  $F_2(1, 35) = 3.55$ ,  $p < 0.07$ ). There was therefore a tendency for participants to be less accurate in their comprehension of anomalous sentences, even though accuracy rates were very high in all cases (CON-conditions: 96.89%; REV-conditions: 94.39%).

### 3.4.2. ERP data

Grand average ERPs time-locked to the onset of the critical clause-final verb are shown in Fig. 4.

Visual inspection of Fig. 4 indicates that, for the BEI-conditions, the reversal anomaly engendered a pronounced centro-parietal negativity in comparison to the control condition. This effect is apparent between approximately 150 and 700 ms post verb onset. For the BA-conditions, the pattern is less clear, though there appears to be a slight frontal negativity for the reversal anomaly in comparison to the control. In accordance with the visual inspection of the data, statistical analyses were carried out for the 150–700 ms time window.

The analysis of the lateral electrodes showed a main effect of ANOMALY ( $F(1, 24) = 11.55$ ,  $p < 0.01$ ), a marginally significant interaction ROI  $\times$  COVERB ( $F(3, 72) = 2.96$ ,  $p < 0.06$ ) and a significant interaction ROI  $\times$  COVERB  $\times$  ANOMALY ( $F(3, 72) = 4.30$ ,  $p < 0.03$ ). Resolving the three-way interaction by ROI showed interactions of COVERB  $\times$  ANOMALY in both posterior regions (left:  $F(1, 24) = 7.29$ ,  $p < 0.02$ ; right:  $F(1, 24) = 4.36$ ,  $p < 0.05$ ). A further resolution by COVERB for the regions showing the interaction revealed significant effects of ANOMALY for the BEI-conditions (left-posterior:  $F(1, 24) = 43.46$ ,  $p < 0.0001$ ; right-posterior:  $F(1, 24) = 49.48$ ,  $p < 0.0001$ ), but only a marginal right-posterior effect for the BA-conditions (left-posterior:  $F(1, 24) < 1$ ; right-posterior:  $F(1, 24) = 3.64$ ,  $p < 0.07$ ).

The analysis of the midline electrodes revealed similar results. The global analysis showed a main effect of ANOMALY ( $F(1, 24) = 29.62$ ,  $p < 0.0001$ ) and interactions COVERB  $\times$  ANOMALY ( $F(1, 24) = 4.27$ ,  $p < 0.05$ ) and ROI  $\times$  COVERB  $\times$  ANOMALY ( $F(5, 120) = 5.49$ ,  $p < 0.01$ ). Interactions of COVERB  $\times$  ANOMALY were observable at electrodes Cz ( $F(1, 24) = 6.73$ ,  $p < 0.02$ ), CPz ( $F(1, 24) = 27.43$ ,  $p < 0.0001$ ) and Pz ( $F(1, 24) = 16.12$ ,  $p < 0.001$ ). A subsequent resolution of these interactions by COVERB consistently revealed effects of ANOMALY for the BEI-conditions (all  $F_s > 43$ , all  $p_s < 0.0001$ ) but not for the BA-conditions (all  $F_s < 2.4$ , all  $p_s > 0.13$ ).

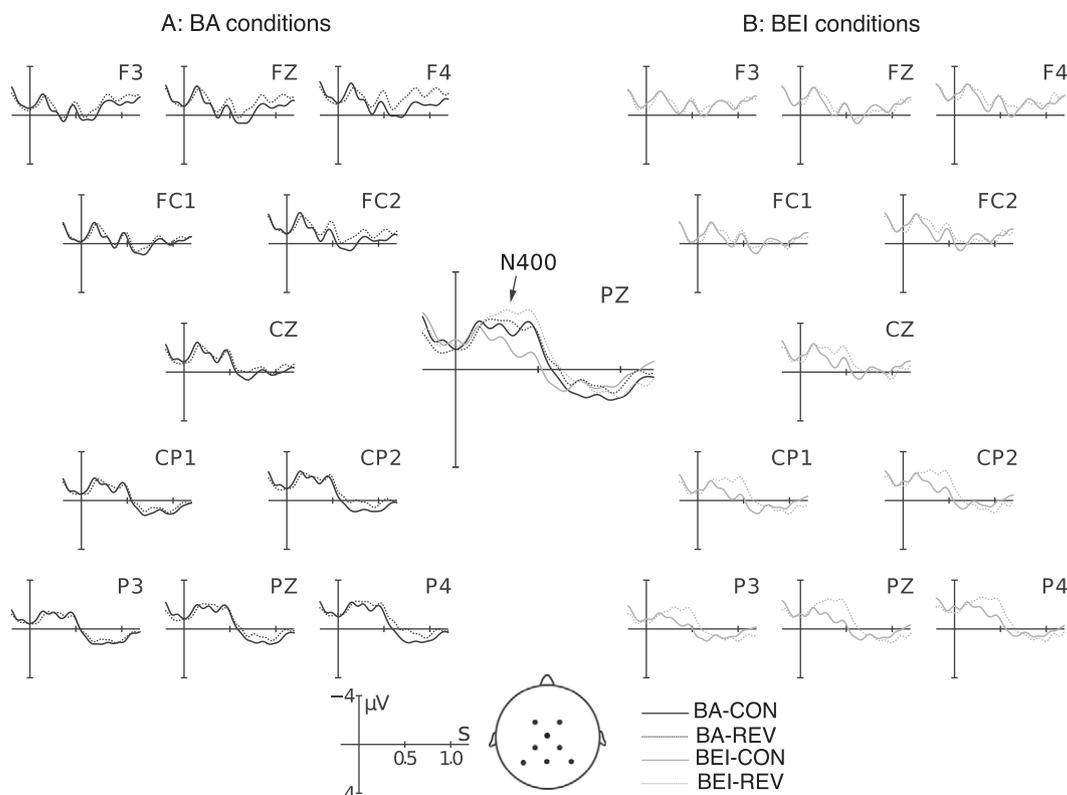


Fig. 4. Grand-average ERPs time-locked to the critical clause-final verb (onset at the vertical bar) in Experiment 2 (Mandarin Chinese). Negativity is plotted upwards.

In summary, the Chinese data from Experiment 2 showed a clear N400 for reversal anomalies vs. controls in *bèi*-sentences, but only a marginal negativity for reversal anomalies in *bā*-structures.

#### 4. Interim discussion

Experiments 1 and 2 revealed that semantic reversal anomalies in Turkish and Mandarin Chinese correlate with N400 effects. This finding demonstrates that the N400 observed for German by [Schlesewsky and Bornkessel-Schlesewsky \(2009\)](#) was not simply an exceptional result. Rather, the overall data pattern suggests that the presence or absence of an N400 in reversal anomalies is subject to cross-linguistic variation: N400 absent (English/Dutch) vs. N400 present (German/Turkish/Chinese). The distribution of the P600/late positivity appears to be somewhat more complex: whereas our original hypothesis was that the late positivity observed following the N400 for reversal anomalies in German reflects a well-formedness mismatch ([Schlesewsky & Bornkessel-Schlesewsky, 2009](#)), the absence of a positivity in Turkish and Chinese appears to call this interpretation into question. We shall return to this issue in the General Discussion; in the remainder of this interim discussion, by contrast, we will focus on the cross-linguistic variation observed for the N400.

Strikingly, the variation in the presence/absence of an N400 across languages does not appear to be governed by language family or by the “perceived” similarity between languages, as German should be predicted to pattern with English and Dutch rather than with Chinese and Turkish under both of these criteria. Rather, the cross-linguistic differences appear to correlate with the properties of the form-to-meaning mapping in a given language: whereas verb-argument linking in English and Dutch is primarily determined by the linear position of the arguments, the flexible ordering

properties of German, Turkish and Chinese require that multiple information types be taken into account during linking. From the perspective of the eADM, this simple dichotomy with regard to linking characteristics has profound consequences for the processing of a reversal anomaly: in languages of the English type, a reversal anomaly does not give rise to a linking problem because linking is vacuously satisfied by the fact that one argument precedes the other; in languages of the German type, by contrast, a reversal anomaly disturbs the linking process because this process must necessarily consider a diverse set of information types (including animacy). Thus, the presence vs. absence of an N400 for reversal anomalies in German, Turkish and Chinese on the one hand and English and Dutch on the other can be modelled as resulting from the presence vs. absence of a conflict in the linking step of the eADM (cf. [Bornkessel-Schlesewsky & Schlesewsky, 2008](#); [Bornkessel-Schlesewsky & Schlesewsky, 2009b](#)).

Crucially, the fact that Chinese patterns with Turkish and German rather than with English and Dutch provides strong converging support for the assumption that word order flexibility is a key feature of the more flexible linking pattern that appears to correlate with the N400 for reversal anomalies. An alternative scenario based on the morphological richness of a language is ruled out by the Chinese results, since this language has very little morphology but nevertheless shows a similar result to German and Turkish (languages with morphological case marking) rather than English and Dutch (languages with very little or no morphological case marking).

In spite of the rather consistent overall pattern of results, a somewhat unexpected finding was that, in Experiment 2, a reliable electrophysiological effect for semantic reversal anomalies vs. controls was observed only for the *bèi*- but not for the *bā*-structures. Could this be in some way related to the fact that, in the *bā*-sentences, the implausible reading (BA-REV) nevertheless conforms to the animacy hierarchy (i.e. the Actor is animate and the

Undergoer is inanimate)? An account along these lines appears unlikely in view of the Turkish findings from Experiment 1, in which we observed reliable N400 effects for both the AI and IA conditions (i.e. regardless of the relative animacy status of the arguments). While the N400 was slightly less pronounced for condition IA–REV at parietal midline sites (i.e. for the condition corresponding to condition BA–REV in Chinese in that the sentence structure called for the animate first NP as the Actor, while the verb semantics designated this argument as the Undergoer), this is in no way comparable to the almost complete absence of an effect for the bǎ-structures in Experiment 2. A possible alternative explanation for the Chinese results is that the processing system is not yet forced to adopt the anomalous reading at the verb position in the bǎ-sentences because they readily permit a continuation as a relative clause. This possibility, which occurs because Chinese relative clauses are prenominal, is illustrated in example (7).

- 
- (7) 侦探 把 [子弹 击中 的] 罐头 拿走了。  
 zhēntàn bǎ zǐdàn jīzhòng de guàntou nǎzǒu-le  
 detective BA bullet hit DE tin take-away-ASP  
 ‘The detective took away the tin which the bullet hit.’
- 

For the bèi-sentences, by contrast, a relative clause structure as in (7) is less likely because a string such as *detective bei bullet keep* would require *bullet* to be the Actor of *keep* even if *keep* were the verb within the relative clause. This possibility therefore does not help to resolve the reversal anomaly.<sup>5</sup> In summary, while the reversal anomaly could be detected relative to the onset of the verb for the bèi-structures, this was not possible for the bǎ-structures due to the possibility of a relative clause analysis. Hence, the detection of the reversal anomaly likely only occurred after sentence offset in the bǎ-conditions, when it became clear that the sentence was a simple main clause. This assumption is corroborated by the acceptability data, which show that the detection rate for the anomaly is lower in the bǎ-structures than in the bèi-structures. The striking difference between the bǎ- and bèi-sentences in Experiment 2 thus appears to be attributable to the interplay between verb-based constraints and the availability of a relative clause reading rather than to principled differences between bǎ- and bèi-constructions.

To sum up this interim discussion, the results from Experiments 1 and 2 provide converging support for the assumptions that (a) there are systematic cross-linguistic differences in the electrophysiological correlates of reversal anomaly processing; and (b) the presence vs. absence of an N400 for these anomalies is tied to the linking properties of the language in question. However, since a crucial characteristic of verb–argument linking is that it is, in part, dependent on the semantic properties of the verb involved (see the introduction), this makes the prediction that it may be possible to change the electrophysiological processing signature for a reversal anomaly by manipulating verb-specific linking properties. Specifically, if, in a particular language, certain verbs particularly emphasise the importance of positional information while others do not, this may manifest itself in a switch between the presence vs. absence of an N400 even with a single language. This prediction was tested in Experiment 3 using Icelandic.

<sup>5</sup> For the bèi-structures, a relative clause reading would only be possible with a subject relative clause such as (i). However, as is apparent from the example, this would require the relative clause verb to be positioned in front of the second noun phrase and is thus not compatible with the sentences used in Experiment 2.

- (i) 侦探 被 [保存 子弹的] 管家 杀死了。  
 zhēntàn bèi bǎocún zǐdàn de guǎnjiā shā-sǐ-le  
 detective BEI keep bullet DE housekeeper kill-dead-ASP  
 ‘The detective was killed by the housekeeper who kept bullets.’

## 5. Experiment 3: Icelandic

### 5.1. Relevant properties of Icelandic

Like English, Dutch and German, Icelandic is a member of the Germanic subbranch of the Indo-European language family. It is similar to English in that it has a basic SVO word order and similar to German and Dutch in that it is a verb-second language, i.e. the finite verb is always the second constituent in a declarative main clause. In contrast to all of the languages in which semantic reversal anomalies have been examined to date, Icelandic has both a rich morphological case marking system and a relatively rigid word order. However, word order is not completely fixed, since object topicalisation to the prefield (i.e. the position preceding the finite verb in main clauses) is possible. This is illustrated in example (8), from Rögvaldsson and Thráinsson (1990, p. 3).

- 
- (8) a. Ég hef aldrei hitt Maríu.  
 I.NOM have never met Mary.ACC  
 ‘I have never met Mary.’  
 b. Maríu hef ég aldrei hitt.  
 Mary.ACC have I.NOM never met  
 ‘I have never met Mary.’
- 

As example (8) shows, the clause-initial (prefield) position is not informative with regard to an argument’s role in Icelandic. There is, however, a fixed subject position which is right-adjacent to the finite verb in a main clause (i.e. the position of *I* in 8b): an argument encountered in this position must necessarily be the subject of the sentence (though the subject need not necessarily be positioned there, cf. 8a). As will become clear below, this property is highly relevant for the design of Experiment 3.

A final crucial characteristic of Icelandic is that the case marking borne by the sentential arguments varies depending on the verb. While most verbs call for a nominative subject and an accusative object (see example 8), dative objects are also possible, as are dative subjects. Dative subjects are classified as “subjects” even though they do not bear nominative case and often show no (or only reduced) agreement with the verb as they display a number of other subject properties e.g. with respect to control, reflexivisation, deletion under conditions of coreference, etc. (for a cross-linguistic overview of non-nominative subjects, see Bhaskararao & Subbarao, 2004). Since case marking patterns are verb-dependent, the relevant properties of verb–argument linking (i.e. which argument should be assigned which role) is typically determined by the combination of case and verb class. There is, however, one exception to this general rule, namely the class of so-called “alternating” verbs (e.g. *to follow*). With these verbs, subject and object can bear nominative and dative case, respectively, but the reversed pattern (dative subject, nominative object) is also possible (Barddal, 2001). For alternating verbs, case is therefore not informative with regard to verb–argument-linking. Rather, role assignment is determined by which argument is placed in the subject position; this is illustrated in (9).

- 
- (9) Examples of an alternating verb in Icelandic (Barddal, 2001, p. 47)  
 a. Hentar þetta þér?  
 pleases this.NOM you.DAT  
 ‘Does this please you?’  
 b. Hentar þér þetta?  
 pleases you.DAT this.NOM  
 ‘Are you pleased by this?’
-

Note that the translations of the two examples in (9) are only approximate: the choice of subject (nominative or dative) does not imply a change in voice. Crucially, however, the alternation between (9a) and (9b) involves a change in role assignments: in (9a), the nominative argument *this* causes a psychological change in the dative argument *you* (i.e. the nominative argument is the Actor); (9b), by contrast, implies that the dative argument is in a state of experiencing something and directs its attention towards the stimulus (i.e. the nominative argument is the Undergoer) (see Van Valin, 1991). Barddal (2001) likens this alternative way of construing the same event to the alternation between *frighten* and *fear* in English: whereas *The darkness frightens Mary* is causative, *Mary fears the darkness* is typically considered stative; hence, only with *frighten*, the stimulus (*darkness*) bears the higher-ranking role (Grimshaw, 1990). Thus, with alternating verbs, argument position determines generalised role assignments.

In summary, the information types relevant to verb-argument linking in Icelandic vary depending on verb type. With most verb classes, role assignments are determined by case marking information. For the so-called alternating verbs, however, role assignments depend on which argument occupies the subject position. Thus, for non-alternating verbs, Icelandic resembles German with regard to its linking characteristics; for alternating verbs, it more closely resembles English. Experiment 3 capitalised upon this distinction in order to examine whether these verb-based linking differences engender qualitatively different ERP responses.

## 5.2. Experimental design and hypotheses

The four critical conditions employed in Experiment 3 are shown in Table 3.

As in Experiments 1 and 2, the critical conditions in Experiment 3 served to compare semantic reversal anomalies (REV-conditions) with their respective controls (CON-conditions). In contrast to the previous studies, Experiment 3 did not use verb-final sentences: due to the basic SVO order of Icelandic, verb-final sentences are only possible in object relative clauses (as in English). Rather, the reversal anomalies were induced via the combination of the subject (placed in the subject position following the adverb) and verb.

The second crucial manipulation in Experiment 3 was one of verb type (non-alternating verbs, NON vs. alternating verbs ALT). As described in the preceding section, these two verb classes differ with regard to their linking properties. Whereas, with non-alternating verbs, case marking is the relevant property for role identification, this is only possible via position for the alternating verbs. By examining reversal anomalies with these different verb types, we therefore aimed to investigate whether the different

linking characteristics correlate with distinct electrophysiological processing signatures.

Our hypotheses were as follows. If we are correct in assuming that different verb-based linking properties can change the way in which semantic reversal anomalies are processed, we should observe an N400 for reversal anomalies with non-alternating verbs (i.e. the pattern observed in German, Turkish and Chinese) and no N400 for reversal anomalies with alternating verbs (i.e. the pattern observed in English and Dutch). In accordance with the pattern from German, we expected to observe a late positivity for all reversal anomalies due to the irresolvability of the conflict. In view of the fact that Turkish and Chinese did not show a positivity, however, this prediction appears less certain than that for the N400 alternation.

## 5.3. Materials and methods

### 5.3.1. Participants

Twenty native speakers of Icelandic (all students from the University of Iceland, Reykjavik; 12 women; mean age: 24.3 years, range: 20–34 years) participated in Experiment 3 after giving written informed consent. All were right-handed and had normal or corrected-to-normal vision. An additional two participants were excluded from the final data analysis due to excessive EEG artefacts.

### 5.3.2. Materials

Twenty sets of the four critical conditions in Table 3 were conducted, thus yielding a total of 80 critical sentences. These were pseudo-randomly interspersed with 200 fillers, 160 of which implemented a separate experimental manipulation and 40 of which served to ensure that inanimate (nominative and dative) subjects did not always lead to an unacceptable (or implausible) sentence.

### 5.3.3. Procedure

Participants were seated at a distance of approximately 1 m in front of a 14 in. laptop screen. Each trial began with the presentation of a fixation asterisk in the centre of the screen (1000 ms), after which the sentence was presented in word-by-word manner. Each word was presented for 750 ms with an inter-stimulus interval (ISI) of 150 ms. This relatively long presentation time was chosen because of the morphological complexity of Icelandic (for similar arguments for Turkish, see Demiral, Schlesewsky, & Bornkessel-Schlesewsky, 2008) and was perceived as a comfortable reading rate by participants. After the presentation of the sentence, there was a 400 ms pause before participants were required to complete an acceptability judgement task (signalled through the

**Table 3**  
Sentence examples and mean acceptability ratings for the four critical conditions in Experiment 3 (Icelandic). Standard deviations (by participants) are given in parentheses. Abbreviations: NON – condition with a non-alternating verb (nominative NP = Actor); ALT – condition with an alternating verb (NP in subject position = Actor); CON – control condition; REV – reversal anomaly.

| Condition | Example  | Acceptability (%) |
|-----------|--|-------------------|
| NON-CON   | Líklega hefur konan <u>treyst</u> hjólinu fullkomlega<br>likely has woman.NOM trusted bicycle.DAT completely<br>'The woman likely trusted the bicycle completely.'   | 91.2 (19.1)       |
| NON-REV   | Líklega hefur hjólið <u>treyst</u> konunni fullkomlega<br>likely has bicycle.NOM trusted woman.DAT completely<br>'#The bicycle likely trusted the woman completely.' | 7.6 (7.4)         |
| ALT-CON   | Yfirleitt hefur manningum <u>farið</u> skeggið vel<br>in.general has man.DAT gone moustache.NOM well<br>'In general, the moustache looked good on the man.'          | 99.5 (1.5)        |
| ALT-REV   | Yfirleitt hefur skegginu <u>farið</u> maðurinn vel<br>in.general has moustache.DAT gone man.NOM well<br>'#In general, the man looked good on the moustache.'         | 10.2 (12.1)       |

presentation of a question mark), which involved judging whether the sentence was acceptable or not. Participants responded by pressing the left or right mouse button for ‘yes’ or ‘no’ (maximal reaction time: 3000 ms). Subsequently, participants completed a probe detection task in which they judged whether a single word had occurred in the previous sentence or not (maximal reaction time: 3000 ms). Trials were separated by an inter-trial interval (ITI) of 1250 ms. Participants were asked to avoid movements and eye blinks during the presentation of the sentences.

Following a short practice session, the experimental sentences were presented in 8 blocks of 35 sentences each. Between blocks, participants took short breaks. Including electrode application, an entire experimental session lasted approximately 2.5 h.

### 5.3.4. EEG recording and preprocessing

The EEG recording and preprocessing procedures were identical to those employed for Experiments 1 and 2 with three exceptions: the data were recorded using a BrainAmp DC amplifier (Brain Products GmbH, Gilching, Germany) and EasyCap electrode caps (EasyCap GmbH, Herrsching, Germany). The sampling rate was 500 Hz.

### 5.3.5. Data analysis

The data analysis was analogous to that in Experiments 1 and 2 with the exception that Experiment 3 included the two condition factors ANOMALY (REV: reversal anomaly vs. CON: control) and VERB (NON: non-alternating verb vs. ALT: alternating verb).

## 5.4. Results

### 5.4.1. Behavioural data

As in Experiments 1 and 2, participants showed a high degree of accuracy in performing the behavioural tasks (mean accuracy rates per condition: >89% for the judgement task, see also Table 3; >96% for the probe detection task).

The analysis of the accuracy rates for the acceptability judgement task revealed an interaction ANOMALY  $\times$  VERB ( $F_1(1, 19) = 4.01, p < 0.06; F_2(1, 19) = 7.78, p < 0.02$ ). Resolving the interaction

by VERB revealed an accuracy difference between the CON and REV-conditions for the alternating verbs ( $F_1(1, 19) = 11.921, p < 0.01; F_2(1, 19) = 24.42, p < 0.0001$ ) but not for the non-alternating verbs ( $F_1/F_2 < 1$ ).

The analysis of the accuracy rates for the probe detection task did not show any significant main effects or interactions.

### 5.4.2. EEG data

Grand average ERPs at the position of the critical verb are shown in Fig. 5.

As is apparent from Fig. 5, the pattern for reversal anomaly processing appears to differ depending on verb type. For the non-alternating verbs, reversal anomalies vs. controls engendered a central negativity between approximately 400 and 600 ms followed by a late positivity (600–800 ms). For the alternating verbs, by contrast, the reversal anomalies elicited only a late positivity effect (600–800 ms). In accordance with these descriptive impressions, statistical analyses were conducted in two time windows: 400–600 ms for the N400 effect and 600–800 ms for the late positivity.

In the 400–600 ms time window, the analysis of the lateral electrode sites showed interactions of ROI  $\times$  ANOMALY ( $F(3, 57) = 7.42, p < 0.0001$ ) and VERB  $\times$  ANOMALY ( $F(1, 19) = 4.93, p < 0.04$ ). Resolving the latter by VERB revealed an effect of ANOMALY only for the non-alternating verbs ( $F(1, 19) = 7.23, p < 0.02$ ) but not for the alternating verbs ( $F < 1$ ). For the midline electrodes, we also observed an interaction of ROI  $\times$  ANOMALY ( $F(5, 95) = 8.35, p < 0.01$ ) and a marginal interaction of VERB  $\times$  ANOMALY ( $F(1, 19) = 3.27, p = 0.08$ ). The VERB  $\times$  ANOMALY interaction was again due to a significant effect of ANOMALY for the non-alternating verbs ( $F(1, 19) = 5.45, p < 0.04$ ) but not for the alternating verbs ( $F < 1$ ).

In the 600–800 ms time window, the statistical analysis revealed a main effect of ANOMALY (lateral:  $F(1, 19) = 4.96, p < 0.04$ ; midline:  $F(1, 19) = 5.12, p < 0.04$ ) and, for the midline electrodes, an interaction of ROI  $\times$  ANOMALY ( $F(5, 95) = 7.11, p < 0.01$ ). Resolving the interaction by ROI showed effects of ANOMALY at electrodes Fz

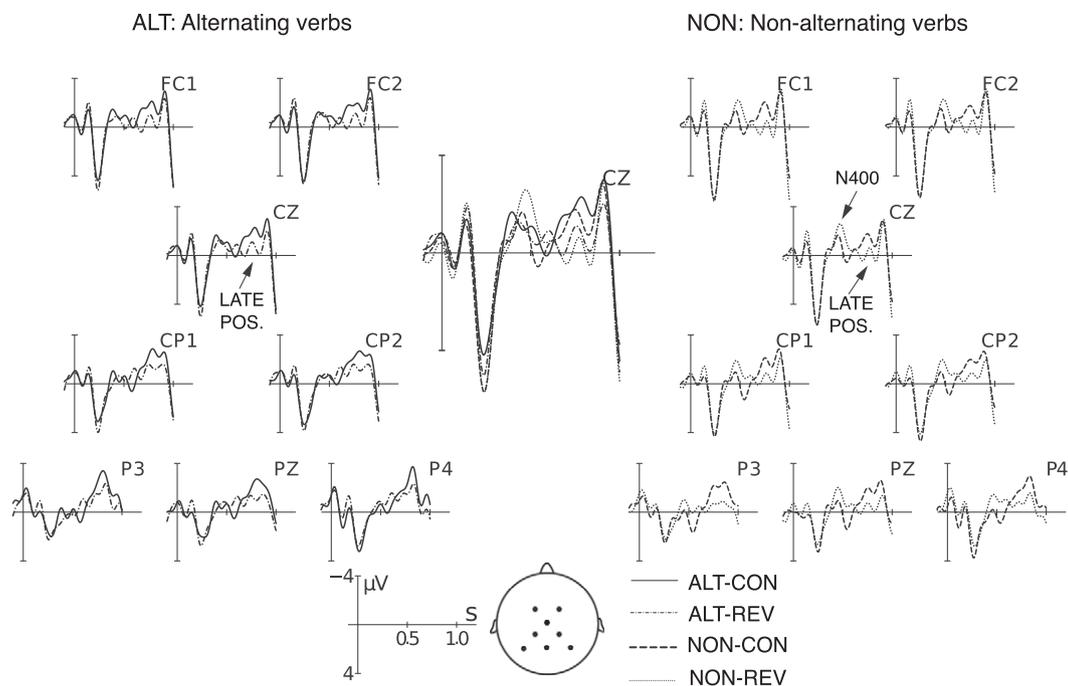


Fig. 5. Grand-average ERPs to the critical verb (onset at the vertical bar) in Experiment 3 (Icelandic). Negativity is plotted upwards.

( $F(1, 19) = 19.68, p < 0.001$ ), FCz ( $F(1, 19) = 9.21, p < 0.01$ ) and Cz ( $F(1, 19) = 4.18, p < 0.06$ ).

In summary, the results of Experiment 3 showed that, for alternating verbs, reversal anomalies show an N400 – late positivity pattern in comparison to controls. By contrast, reversal anomalies in sentences with non-alternating verbs only engendered a late positivity effect in comparison to their respective controls. Interestingly, the late positivity in this experiment had a fronto-central maximum.

## 6. General discussion

We have presented three ERP experiments on the processing of semantic reversal anomalies across languages. Experiments 1 and 2 showed that animacy-based reversal anomalies correlate with N400 effects in Turkish and Mandarin Chinese. Experiment 3 further revealed a verb class dependent pattern for Icelandic: with alternating verbs, in which argument role assignments are determined positionally, we observed a monophasic P600 effect for reversal anomalies versus controls; non-alternating verbs requiring case-based role assignments, by contrast, showed a biphasic N400 – late positivity pattern for reversal anomalies versus controls.

When considered together with previous findings from German (Schlesewsky & Bornkessel-Schlesewsky, 2009), English (Kim & Osterhout, 2005; Kuperberg, Sitnikova et al., 2003; Kuperberg, Caplan et al., 2006; Kuperberg, Kreher et al., 2007) and Dutch (Hoeks et al., 2004; Kolk et al., 2003; van Herten, Kolk et al., 2005; van Herten, Chwilla et al., 2006), the present results reveal a cross-linguistic pattern of electrophysiological processing signatures for reversal anomalies that is summarised in Table 4, together with relevant properties of the individual languages under consideration.

As is apparent from Table 4, the cross-linguistic results confirmed our prediction of a dichotomy between languages showing an N400 and languages showing no N400 for reversal anomalies. With regard to the late positivity, by contrast, the overall pattern turned out to be somewhat more complex than we had anticipated. While the results from German indicated that the irresolvability of a reversal anomaly engenders a late positivity in addition to an N400 (Schlesewsky & Bornkessel-Schlesewsky, 2009) and this pattern was replicated by the results for the non-alternating verbs in Icelandic (Experiment 3), reversal anomalies correlated with a monophasic N400 in Turkish and Chinese in spite of their irresolvability (Experiments 1 and 2). This suggests that there are, in fact, two dimensions of cross-linguistic variation at work here, one affecting the N400 and one affecting the late positivity. These two aspects will be discussed in turn in the following.

### 6.1. Sequence dependence and cross-linguistic differences in the N400

As already discussed in detail in the interim discussion (Section 4), the presence or absence of an N400 for reversal anomalies cuts across language families (English, Dutch, German and Icelandic all belong to the Germanic branch of the Indo-European language family, while Turkish is an Altaic and Mandarin Chinese is a Sino-Tibetan language) and across a range of typological characteristics (e.g. verb-final vs. verb-medial basic word order; rich vs. impoverished case morphology; see Table 4). Rather, the crucial factor in determining whether a language falls into one or the other category appears to be related to word order properties: whereas English and Dutch have rather rigid word orders, German, Turkish and Chinese do not. The results from Icelandic provide converging support for this proposal by showing an “English-style” pattern (i.e. no N400 effect) for verbs requiring position-based argument role identification and a “German-style” pattern (i.e. an N400 effect) for verbs with which role assignments are determined via case marking. These verb-dependent electrophysiological processing signatures further indicate that verb-based linking properties play a crucial role in determining the observed dichotomy between the presence and absence of N400 effects. By contrast, construction-specific but verb-independent properties (e.g. word order freezing in non-case-marked sentences in Turkish) appear less relevant in this regard.

Whether reversal anomalies in a given language engender an N400 effect or not thus appears to correlate with the degree of sequence-dependence of the form-to-meaning mapping in that language. Whereas sequence-independence is reflected in N400 effects (as in German, Turkish, Chinese and with non-alternating verbs in Icelandic), no N400 effects are observed in the case of a strong sequence-dependency.

Interestingly, the cross-linguistic distinction observed here is compatible with a range of further findings in the electrophysiological literature on language comprehension. Thus, while grammatical conflicts in English (and Dutch) have – with the notable exception of the semantic P600 literature – long been associated with P600 effects (Hagoort, 2003; cf. Kutas, Van Petten, & Kluender, 2006; Osterhout et al., 2004), a range of recent findings have provided evidence for “rule-based” N400 effects in other languages (e.g. German: Bornkessel, McElree, Schlesewsky, & Friederici, 2004; Frisch & Schlesewsky, 2001, 2005; Haupt et al., 2008; Janssen, Wiese, & Schlesewsky, 2005; Chinese: Philipp, Bornkessel-Schlesewsky, Bisang, & Schlesewsky, 2008; Wang, Schlesewsky, Bickel, & Bornkessel-Schlesewsky, 2009; Japanese: Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008; Hindi: Choudhary, Schlesewsky, Roehm, & Bornkessel-Schlesewsky, 2009). Notably, the languages for which these effects have been

**Table 4**  
Summary of the cross-linguistic ERP results for the processing of reversal anomalies and relevant properties of the languages examined. For case morphology, classifications are taken from the World Atlas of Language Structures Online (Haspelmath, Dryer, Gil, & Comrie, 2008): the number of cases is given according to Iggesen (2008a); in addition, in contrast to German, Icelandic and Turkish, English and Dutch are classified as languages in which only a particular minority noun phrase type (pronouns) shows case marking, while Mandarin Chinese is classified as having no case marking at all (Iggesen, 2008b). The ranking of cues according to the Competition Model is adapted from Bates et al. (2001) and shows the relative importance of word order and animacy in assigning the actor role across languages (as assessed via offline tasks). Abbreviations: anim = animacy; SVO = subject–verb–object; SOV = subject–object–verb; RCs = relative clauses; WO = word order.

| Language  | ERP pattern |          | Case morphology (no. cases) | Basic word order | Cue ranking (competition model) | Sequence-dependent linking | Prenominal RCs |
|-----------|-------------|----------|-----------------------------|------------------|---------------------------------|----------------------------|----------------|
|           | N400        | Late pos |                             |                  |                                 |                            |                |
| English   | –           | +        | ≤2                          | SVO              | WO>anim                         | +                          | –              |
| Dutch     | –           | +        | ≤2                          | SVO/SOV          | WO>anim                         | +                          | –              |
| Icelandic | ±           | +        | ≥4                          | SVO              | ?                               | ±                          | –              |
| German    | +           | +        | ≥4                          | SVO/SOV          | anim>WO                         | –                          | –              |
| Turkish   | +           | –        | ≥4                          | SOV              | anim>WO                         | –                          | +              |
| Mandarin  |             |          |                             |                  |                                 |                            |                |
| Chinese   | +           | –        | ≤2                          | SVO              | anim>WO                         | –                          | +              |

demonstrated (e.g. for the processing of word order variations or case marking conflicts) all fall into the category of “sequence-independent” languages as defined above. Thus, while the correlation is of course not perfect, there does appear to be a general tendency for these types of languages to more readily show grammatically relevant N400 effects than is the case in sequence-dependent languages such as English.<sup>6</sup>

What, then, is the relationship between the grammatically relevant N400 in sequence-independent languages and the lexical-semantic N400, which appears to occur reliably across languages (Friederici, Pfeifer, & Hahne, 1993, for German; cf. Nakagome et al., 2001, for Japanese; Ye et al., 2006, for Chinese)? While there has not been a lot of research on this subject to date, Roehm, Bornkessel-Schlesewsky, and Schlewsky (2007a) concluded that, in spite of their superficial similarity, there are differences between the two types of N400s. Using time–frequency analyses, these authors compared a word order-related (and hence grammatically relevant) N400 effect with a classical lexical-semantic N400 effect in German and observed differences with regard to (a) the frequency bands in which the N400-related power changes were observable, and (b) the mechanism responsible for the overall change in evoked power (amplitude increases vs. more consistent phase locking across trials). Thus, at least for German, existing findings suggest that lexical-semantic and “grammatical” N400s do not constitute qualitatively identical neurophysiological processing correlates. Nevertheless, they do of course show similarities, perhaps due to partial overlaps in the widely distributed neural networks that are presumably involved in generating the N400 (see Maess, Herrmann, Hahne, Nakamura, & Friederici, 2006, for MEG evidence). It was this overall situation of partially but not completely identical N400 effects that led Haupt et al. (2008) to propose that, akin to proposals for components such as the P300, there may be an “N400 family” rather than a monolithic N400 effect.

Taking the proposal of an N400 family seriously, the common denominator between the N400 effects observed here for semantic reversal anomalies in sequence-independent languages and the more commonly known lexical-semantic N400 effects may lie in the sequence-independent combination of meaningful elements. Lexical associations and semantic representations are, by definition, sequence-independent. This property is shared by processes of verb–argument linking in sequence-independent languages, but contrasts with the sequence-dependent linking mechanisms required in a language such as English.

## 6.2. Cross-linguistic variation in the late positivity: Challenges for a functional dissociation between the monophasic and biphasic patterns

Within the eADM, the original interpretation of monophasic P600 effects for reversal anomalies in English and Dutch was in terms of a conflict within the generalised mapping step, which arises when the role assignments undertaken by compute linking are incompatible with the output of plausibility processing (i.e. the lexical-semantic associations between the argument(s) and the verb) (see Section 1.2 for details). For late positivities following N400 effects in German, by contrast, the positivity was attributed to the failure of a final well-formedness check

<sup>6</sup> The general dissociation between sequence-dependent and sequence-independent linking mechanisms in different languages raises the possibility that so-called local coherence phenomena (i.e. the tendency for irrelevant interpretations to interfere with comprehension in sentences such as *The coach smiled at the player tossed a frisbee*, see Tabor, Galantucci, & Richardson, 2004) may be limited to (or at least considerably stronger in) languages such as English, since the positional adjacency of the critical constituents together with the language's high weighting of positional information biases towards the irrelevant reading. The authors thank Dan Jurafsky for pointing out the potential relevance of these phenomena.

(Schlewsky & Bornkessel-Schlewsky, 2009). From this perspective, the positivities observed in the monophasic pattern are considered functionally distinct from those forming part of the biphasic pattern.<sup>7</sup>

The present experiments challenge this interpretation in several respects. Firstly, while Turkish and Chinese patterned with German in showing an N400 effect for reversal anomalies, the N400 was not followed by a late positivity in these two languages. This is clearly problematic for an interpretation of the late positivity in terms of well-formedness, since the reversal anomalies employed in Turkish and Chinese were reliably judged as unacceptable (showing an acceptability of <3% in Turkish and <13% in Chinese). Notably, acceptability ratings for these sentences (especially in the case of Turkish) were lower than that for the German reversal anomalies examined by Schlewsky and Bornkessel-Schlewsky (2009), which were judged to be acceptable in approximately 18% of all cases. Thus, a violation of well-formedness requirements does not straightforwardly account for the cross-linguistic distribution of the late positivity. Secondly, recall that reversal anomalies in Icelandic showed either a monophasic P600 effect or a biphasic N400 – late positivity pattern depending on verb type. In terms of the explanation outlined above, the two positivities would receive functionally distinct interpretations (in terms of generalised mapping and well-formedness, respectively). However, the data do not provide any empirical support for this model-internal assumption: there was no hint of an interaction between anomaly and verb type in the late positivity time window in Experiment 3. In the following, we will discuss two potential solutions to these problems, before turning to an alternative explanation in the next section.

With regard to the presence of a late positivity in German and Icelandic and its absence in Turkish and Chinese, one possibility appears to be that the distinction depends on the modality of presentation. Whereas Schlewsky and Bornkessel-Schlewsky's (2009) study on German and Experiment 3 on Icelandic used visual presentation, Experiments 1 and 2 on Turkish and Chinese used auditory presentation. It is therefore possible that the presence or absence of a late positivity following the N400 correlates with visual vs. auditory presentation, respectively. If true, this explanation would suggest that the observation of late positivities for reversal anomalies is somehow tied to the use of visual presentation. However, previous results challenge this assumption: Wassenaar and Hagoort (2007) observed a P600 – without an accompanying N400 – for reversal anomalies in Dutch using auditory presentation. While this experiment used a sentence–picture matching task, the similarity between the effects observed and those found in other studies on semantic reversal anomalies in Dutch clearly attests to the fact that reversal anomalies can, in principle, engender late positivities in the auditory modality. It thus appears unlikely that the difference observed between German and Icelandic on the one hand and Turkish and Chinese on the other should be attributed to differences in the modality of presentation.

A second possibility is that the presence vs. absence of a late positivity is due to some property of the languages under consideration (as we also claim for the N400, see Section 6.1). If, as postulated within the eADM, it is indeed the case that a late positivity following an N400 typically correlates with the resolvability of a processing conflict (see Bornkessel-Schlewsky & Schlewsky, 2008; Schlewsky & Bornkessel-Schlewsky, 2009), the availabil-

<sup>7</sup> As noted by Bornkessel-Schlewsky and Schlewsky (2008), the monophasic positivity could result from an overlap of generalised mapping and well-formedness, since reversal anomalies are also irresolvable in English and Dutch. Crucially, however, the late positivity that forms part of the biphasic pattern is not viewed as a correlate of generalised mapping.

ity of prenominal relative clauses in Turkish and Chinese could potentially play a crucial role in this regard. Recall from the interim discussion (Section 4) that, at the position of the verb in the Chinese sentences used here, a biclausal analysis of the input is potentially possible because relative clauses in Chinese are prenominal (i.e. occur before the noun which they modify). This also applies for Turkish, which is a strictly head-final language (e.g. Kornfilt, 1997). Thus, in both Turkish and Chinese, the input string NP–NP–V is structurally ambiguous between a monoclausal reading and a structure such as (10), in which the second NP and the verb form part of a relative clause. The availability of the alternative structural analysis thus, in principle, renders the processing conflict induced by the reversal anomaly resolvable (for evidence that potentially conflicting information within an assumed main clause leads the processing system to adopt a biclausal reading in head-final languages, see Miyamoto, 2002). This is not the case in German and Icelandic, which are neither strictly head-final nor have prenominal relative clauses, and thus do not provide a potential “way out” of the conflict via a biclausal analysis of the input.

---

(10) NP [NP V]<sub>RC</sub> NP V

---

While an account along these lines offers a potentially plausible explanation for the presence vs. absence of a late positivity for reversal anomalies in German/Icelandic vs. Turkish/Chinese, several problems remain. Firstly, as already noted in the interim discussion in Section 4, the possibility of a relative clause reading appears much less likely for the *bèi*- than for the *bā*-structures in Chinese. One would therefore have to claim that resolvability is determined on a purely structural basis, rather than with reference to the specific constraints imposed by the lexical and semantic information of the input items. Secondly, if the well-formedness assessment within the eADM were completely equated with the issue of resolvability (as required by this account), the association between well-formedness and sentence acceptability would need to be relinquished. Finally, while sketching out a possible explanation for why Turkish and Chinese do not show a late positivity for reversal anomalies, this account does not solve the second problem outlined above, the lack of empirical evidence for a distinction between a generalised mapping positivity (the monophasic P600) and a well-formedness positivity (the late positivity that forms part of a biphasic pattern).

In view of these remaining issues, the next section presents a possible alternative account of the late positivity effects for reversal anomalies.

### 6.3. A unified account of late positivities for reversal anomalies as categorisation-related P300s

As discussed in the preceding section, the assumption of a functional distinction between monophasic P600s and late positivities which form part of a biphasic pattern in response to reversal anomalies appears problematic in several respects. In this section, we will therefore consider an alternative possibility, namely that of a unified explanation for all of the late positive effects engendered by semantic reversal anomalies, irrespective of the presence/absence of a preceding N400. Specifically, we will consider the possibility that these positivities are instances of a late, target-related P300, which reflects a binary stimulus categorisation (well-formed vs. ill-formed). In this respect, we draw upon an account of the target-related P300 in language processing that was proposed by Kretzschmar (2010).

The P3b, as one subcomponent of the P300 family, is a positivity with a posterior scalp distribution, which is elicited by task-relevant target stimuli (for reviews, see Donchin, 1981; Pritchard,

1981; Verleger, 1988). Its amplitude increases for salient and improbable stimuli and its latency has been interpreted as an index of stimulus evaluation time (for a review, see Pritchard, 1981; for a slightly different view, see Verleger, 1997). Crucially, the P3b (henceforth: P300) appears to reflect processes of stimulus categorisation (as required, for example, when task-relevant target stimuli need to be dissociated from non-targets in an oddball paradigm) and this categorisation process needs to be based on a binary decision (Donchin & Coles, 1988). Hence, in linguistic contexts, P300 effects are not typically observed in response to semantic incongruities (Kutas & Hillyard, 1980), because the categorisation between an incongruous and a congruous stimulus is not binary (i.e. the incongruity in *He spread the warm bread with socks* contrasts with a number of possible congruous continuations, e.g. *butter, jam, honey* etc.). By contrast, when a semantic incongruity is introduced in sentences in which there is only a single congruous continuation (e.g. *The opposite of black is nice* vs. *The opposite of black is white*), a P300 is observable (Kutas & Iragui, 1998; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlewsky, 2007b), reflecting the binary nature of the decision between the (single) congruous continuation and an incongruous continuation. Finally, the latency of P300 effects engendered by language tasks depends on the processes required for the categorisation of the stimulus: in lexical decision tasks, which require a binary decision (“word” vs. “not a word”), orthographically illegal nonwords engender a P300 within the time range of the N400, while orthographically legal pseudowords engender a P300 following the N400 (Bentin, McCarthy, & Wood, 1985; Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Holcomb & Neville, 1990; Roehm et al., 2007b). In the case of nonwords, the “not a word” categorisation can be undertaken on the basis of prelexical properties (phonology, orthography), whereas the categorisation of a pseudoword requires a lexical search. Hence, the P300 is delayed to the post-N400 time range for pseudowords but not for nonwords.

These characteristics of the P300 readily map onto the late positivity effects observed in the domain of semantic reversal anomalies. Firstly, studies examining reversal anomalies have typically employed judgement tasks, thereby requiring a categorisation of sentences as well-formed vs. ill-formed. Reversal anomalies are thus clearly task-relevant (for an instruction-based modulation of the late positivity for reversal anomalies in Dutch, see Vissers, Chwilla, & Kolk, 2007). Secondly, since reversal anomalies have a lower linguistic a priori probability than their respective control sentences, their subjective probability is lower than that of controls even when the experimental probability of anomalies vs. controls is equated. In view of this reduced degree of subjective probability, reversal anomalies should be expected to show a larger P300 amplitude than control sentences (Donchin, 1981; Donchin & Coles, 1988; Johnson, 1986). Thirdly, since the detection of a reversal anomaly requires lexical processing of the critical word to have been completed, the P300 should be expected to occur in the post-N400 time range. Finally, in contrast to the semantic incongruities that typically engender N400 effects (as first examined by Kutas & Hillyard, 1980), semantic reversal anomalies allow for a binary categorisation, since the anomaly, as its name suggests, involves the reversal of the argument role assignments in the non-anomalous control sentence. Thus, by interpreting the late positivity effect for reversal anomalies as a P300, we can account for its distribution and latency as well as for the fact that it occurs in reversal anomalies as opposed to other types of semantic incongruities.

Perhaps even more importantly, this account also offers a possible explanation for the absence of a late positivity effect for reversal anomalies in Turkish and Chinese. Recall from the previous subsection that, in contrast to all of the other languages in which reversal anomalies have been studied to date, there is a

structural ambiguity at the position of the anomaly-inducing word in Turkish and Chinese because of the possibility of a prenominal relative clause (or, in Turkish, any kind of subordinate clause). While it is difficult to argue convincingly for an impact of this structural ambiguity on conflict resolvability (see Section 6.2), it appears quite possible that it could influence the binarity of the categorisation decision: as soon as the grammar of the language, in principle, permits another structural alternative, the categorisation between “reversal anomaly” vs. “non-anomaly” is no longer a binary choice, since there are at least two possible options for a non-anomalous reading. Since the elicitation of a P300 presupposes a binary choice in categorisation, no positivity effect is engendered by reversal anomalies in Turkish and Chinese.

Converging support for this proposal stems from a recent ERP study on Chinese, which presented NP–V–NP structures such as (11) (Wang et al., 2009):

|      |  |          |         |
|------|--|----------|---------|
| (11) | 小说   | 阅读了      | 演员。     |
|      | xiǎoshuō   | yuèdú-le | yǎnyuán |
|      | novel  | read-ASP | actor   |
|      | *‘The actor read the novel.’/#‘The novel read the actor’ |          |         |

At the position of the second NP (*actor*), sentences such as (11) engendered an N400 – late positivity pattern in comparison to sentences in which the order of NPs was reversed (i.e. *yanyuan yuedu le xiaoshuo*, ‘the actor read the novel’). Whether sentences such as (11) should be considered reversal anomalies is not as straightforward as for the sentences used in the present experiments: while object-verb-subject orders are considered ungrammatical in Chinese (Li & Thompson, 1981), thus speaking in favour of a reversal anomaly interpretation, behavioural studies have shown that native speakers of Chinese tend to interpret the animate argument in these structures as the Actor, thus speaking against a reversal anomaly interpretation. Importantly, however, these sentences clearly bear a strong resemblance to reversal anomalies. The fact that, in addition to an N400 effect, they engender a late positivity effect corroborates the P300 interpretation advanced above: in the postverbal position, a relative clause analysis would no longer help to resolve the anomaly because it would only serve to introduce a complex postverbal NP without changing the processing problem that is induced by *novel-read-NP*. Hence, a binary categorisation is again possible, thus leading to a (late) P300. However, while this interpretation offers a plausible explanation for the range of cross-linguistic findings presented in this paper, it is important to note that it was not predicted in advance and thus clearly requires further corroboration in future research.

In summary, by interpreting the positivity effect for reversal anomalies as an instance of a categorisation-related P300 (P3b), we can provide a unified explanation for the presence vs. absence of this effect across languages: it is always present when the grammar of the language under consideration allows for a binary anomaly vs. non-anomaly decision to be made at the position of the critical word. From this perspective, there are two independent cross-linguistic dimensions of variation in the processing of reversal anomalies: the presence vs. absence of an N400 (conditioned by sequence-dependence, see Section 6.1) and the presence vs. absence of a late P300 (conditioned by the binarity of the categorisation between anomalous and non-anomalous structures).

#### 6.4. Consequences for the neurocognitive modelling of language

Our findings have revealed striking cross-linguistic differences in the electrophysiological correlates of the form-to-meaning mapping during sentence comprehension. We have argued above that

these are most parsimoniously explained via two independent dimensions of variation, which are reflected in the N400 and the late positivity, respectively.

The presence vs. absence of an N400 effect for reversal anomalies correlates with the sequence-independence vs. sequence-dependence of verb-argument linking in a given language. Why should sequencing stand out as the critical defining property in this regard? Perhaps the simplest answer is that a dominant role of sequential information is not surprising given that it is always present in the linguistic input. However, in view of the fact that a sequential order is essentially available “for free”, it appears striking that word order rigidity appears to be the exception rather than the rule in the languages of the world (Steele, 1978). This suggests that, in spite of the convenience of using linear order information, there may be strong competing motivations which bias against strict word order restrictions. For example, a crucial disadvantage in relying entirely upon sequencing for the form-to-meaning mapping is that it only allows for the encoding of binary relations (i.e. given A and B, either A precedes B or B precedes A; A acts on B vs. B acts on A). Non-sequential information types such as case marking, by contrast, not only allow for an encoding of the hierarchical relationship between the arguments (A acts on B; “role assignments”) but also the degree to which the arguments are ideal role fillers (“role prototypicality”). In a German sentence such as *Mir ist die Vase zerbrochen* (I.DAT is [the vase].NOM broken, ‘~The vase broke on me.’), the dative argument *mir* is the causer of the breaking event, but not the voluntary agent of the action being described (the sentence cannot be modified by an adverb describing a volitional action such as *absichtlich*, ‘deliberately’). Thus, the dative allows for the interpretation “A is acting on B, but A is not an optimal Actor”; a nuance which cannot be conveyed via the binary relation of A precedes B. In this way, while sequential information necessarily plays a crucial role during language processing because of the inherently temporal nature of language-based communication, there are good reasons for languages to rely on information sources that go beyond linear order.

With regard to the late positivity effects, we have proposed that the cross-linguistic results are best accounted for by interpreting the positivity as a late P300 which reflects stimulus categorisation. This account contrasts with our previous proposal that the monophasic P600 for reversal anomalies should be viewed as a correlate of the generalised mapping step within the eADM (Bornkessel-Schlesewsky & Schlewsky, 2008; Schlewsky & Bornkessel-Schlesewsky, 2009). Rather, it suggests that the positivities engendered by reversal anomalies reflect a domain-general, binary categorisation of well-formedness. If this proposal is correct, the process reflected in the positivity effect is not the conflict between alternative interpretations (as previously proposed), but the detection of the implausibility which allows for stimulus categorisation. This view is supported by ERP results on the processing of semantic anomalies in which the critical word has a good fit to the context (e.g. 12). “Borderline” anomalies of this type, which are only detected in approximately 50% of all cases, only engender a late positivity effect when they are classified as anomalous (Sanford, Leuthold, Bohan, & Sanford, 2009). This result is perfectly in line with a categorisation-based interpretation of the late positivity effect for semantic reversal anomalies.<sup>8</sup>

<sup>8</sup> Interestingly, like semantic reversal anomalies, borderline anomalies do not engender N400 effects in comparison to plausible controls. This suggests that, in both types of structures, the absence of an N400 might (in addition to the absence of a linking problem, as discussed in detail in this paper) be due to a good contextual fit of the critical word, i.e. to a high degree of lexical-semantic association between the critical word and the sentence and/or discourse context (e.g. Lau et al., 2008; for an application to semantic reversal anomalies, see Stroud, 2008).

## (12) Example from Sanford et al. (2009)

Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year sentence was given to the victim, but this was subsequently appealed.

The proposal that the well-formedness evaluation step of the eADM may be related to the P300 has been put forward previously, e.g. on the basis of the fact that the well-formedness assessment is task-dependent (Bornkessel & Schlewsky, 2006; Bornkessel-Schlewsky & Schlewsky, 2008). In light of the present results, this proposal can be taken one step further: rather than being internal to the language processing architecture, the well-formedness check may reflect a domain-general categorisation processes, with cross-linguistic variation explained by the interplay between language-specific grammatical properties and general constraints on categorisation (e.g. binarity).

Crucially, note that this revised interpretation of positivities correlating with reversal anomalies should not be taken to imply that the architecture of the eADM requires revision. While the present proposal revokes the association between monophasic positivities for reversal anomalies and the generalised mapping step, generalised mapping is clearly required in order to account for other phenomena, e.g. for positivities associated with pragmatic enrichment (Burkhardt, 2006; Burkhardt, 2007; Schumacher, *in press*), which cannot be interpreted as correlating with a binary categorisation process. Similarly, we do not mean to claim that all language-related positivities (P600s) should be interpreted as instances of P300 effects (for discussion of this question, see Coulson, King, & Kutas, 1998; Gunter, Stowe, & Mulder, 1997; Osterhout & Hagoort, 1999; Osterhout, McKinnon, Bersick, & Corey, 1996), but only that late positivities for reversal anomalies lend themselves to a P300 interpretation in view of their cross-linguistic distribution.

This proposal bears a certain resemblance to the conflict monitoring hypothesis put forward by Kolk and colleagues (Kolk et al., 2003; van Herten, Kolk et al., 2005; van Herten, Chwilla et al., 2006; Visser, Chwilla, & Kolk, 2006), according to which conflicts encountered during language processing (e.g. in the case of, but not limited to, semantic reversal anomalies) will lead the processing system to recheck whether the input has been processed correctly. Conflict monitoring is viewed as a domain-general process which correlates with late positive ERP effects and which is thus, to some degree, related to the P300 interpretation proposed here. However, categorisation and conflict monitoring are clearly different in functional terms. Furthermore, like a categorisation-independent account based on well-formedness (see Section 6.2), the conflict monitoring account encounters difficulties when faced with the data from Turkish and Chinese, which did not show a late positivity for clearly unacceptable reversal anomalies.

Finally, the present findings clearly indicate that the dimensions of variation which constrain the distribution of N400 and late positivity effects in the cross-linguistic processing of reversal anomalies are best described at a more general level than in terms of the typical subdivision between semantics and syntax. Rather than indicating the dominance of a plausibility-based heuristic in sentence interpretation, the absence of an N400 for reversal anomalies in languages such as English mirrors the sequence-dependence of linking in these types of languages. Likewise, the presence vs. absence of a late positivity does not indicate the presence vs. absence of a syntactic processing problem, but is rather most plausibly explained in terms of general processes of categorisation.

To summarise, the present findings indicate that reliable functional interpretations of neurobiological patterns within the

language domain require systematic cross-linguistic support. By attempting to interpret findings from individual languages or languages of similar types, we run the risk of generalising language-specific properties (e.g. sequence-dependence in English) to more general functional domains (e.g. syntax). The long-standing discussion of syntax vs. semantics within the cognitive neuroscience of language, which has failed to produce consistent results (cf. Bornkessel-Schlewsky & Schlewsky, 2009a; Bornkessel-Schlewsky & Schlewsky, 2009b) is a case in point: as the present study shows, what appears to be a relatively straightforward distinction in English is not nearly as clear-cut in languages of other types. Therefore, the attempt to use N400 and P600 effects as diagnostics for semantic and syntactic processing in other languages (e.g. Ye et al., 2007) or in second language learning (e.g. Kotz, 2009) is highly problematic. (Note that the same considerations apply to functional-neuroanatomical interpretations of fMRI findings.) Rather, we need to consider more flexible accounts of the form-to-meaning mapping, which subsume both sequence-dependent and sequence-independent language types. The concept of verb-argument linking appears to be a promising candidate in this regard (Bornkessel-Schlewsky & Schlewsky, 2009b); it allows for the language-specific development of optimal information processing strategies (cf. Bates et al., 2001; MacWhinney & Bates, 1989; but also Howes, Lewis, & Vera, 2009), which may result in cognitively and neurobiologically distinct processing mechanisms subserving the same overarching function.

### Acknowledgements

The parts of the research reported here were performed while I.B.S., S.G., D.R., and L.W. were at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. Parts of the research reported here were supported by the Heinz Maier-Leibnitz prize (awarded to I.B.S. by the German Research Foundation) and by a grant from the German Research Foundation (BO 2471/2-1/SCHL 544/2-1). Parts of this research were conducted in collaboration with the Clinic for Audiology and Phoniatry (Prof. Manfred Gross) of the Charité Berlin. We are grateful to Katja Bruening and Isabel Plauth for invaluable assistance in data acquisition.

### References

- Aissen, J. (2003). Differential object marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory*, 21, 435–483.
- Barddal, J. (2001). The perplexity of dat-nom verbs in Icelandic. *Nordic Journal of Linguistics*, 24, 47–70.
- Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, 11, 245–299.
- Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*, 52, 369–396.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343–355.
- Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*, 11, 235–260.
- Bhaskararao, P., & Subbarao, K. V. (Eds.). (2004). *Non-nominative subjects*. Amsterdam: John Benjamins.
- Bisang, W. (1992). *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer: Vergleichende Grammatik im Rahmen der Verbserialisierung und der Grammatikalisierung und der Attraktorpositionen*. Tübingen: Narr.
- Bornkessel-Schlewsky, I., & Schlewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, 59, 55–73.
- Bornkessel-Schlewsky, I., & Schlewsky, M. (2009a). *Processing syntax and morphology: A neurocognitive perspective*. Oxford: Oxford University Press.
- Bornkessel-Schlewsky, I., & Schlewsky, M. (2009b). The role of prominence information in the real time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, 3, 19–58.

- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, 51, 495–522.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113, 787–821.
- Bossong, G. (1985). *Differentielle objektmarkierung in den neuiranischen Sprachen* [=differential object marking in the New Iranian languages]. Tübingen: Narr.
- Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98, 159–168.
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport*, 18, 1851–1854.
- Chang, J.-H. (2003). Event structure and argument linking in Chinese. *Language and Linguistics*, 4, 317–351.
- Chen, E., West, W. C., Waters, G., & Caplan, D. (2006). Determinants of BOLD signal correlates of processing object-extracted relative clauses. *Cortex*, 42, 591–604.
- Choudhary, K. K., Schlesewsky, M., Roehm, D., & Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretively-relevant linguistic rules: Evidence from Hindi. *Neuropsychologia*, 47, 3012–3022.
- 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.
- Demiral, S. B. (2008). *Incremental argument interpretation in Turkish sentence comprehension* (doctoral dissertation ed.). Leipzig: Max Planck Series in Human Cognitive and Brain Sciences.
- Demiral, S. B., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from Turkish. *Cognition*, 106, 484–500.
- Donchin, E. (1981). Surprise! ... Surprise? *Psychophysiology*, 18, 493–513.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11, 357–374.
- Erguvanli, E. E. (1984). *The function of word order in Turkish grammar*. Berkeley/Los Angeles/London: University of California Press.
- 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.
- Frisch, S., & Schlesewsky, M. (2001). The N400 indicates problems of thematic hierarchizing. *Neuroreport*, 12, 3391–3394.
- Frisch, S., & Schlesewsky, M. (2005). The resolution of case conflicts from a neurophysiological perspective. *Cognitive Brain Research*, 25, 484–498.
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25, 149–188.
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660–676.
- Grimshaw, J. B. (1990). *Argument structure*. Cambridge, Mass.: MIT Press.
- Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage*, 20, S18–S29.
- Haspelmath, M., Dryer, M., Gil, D., & Comrie, B. (Eds.) (2008). *The world atlas of language structures online*. Munich: Max Planck Digital Library; <<http://wals.info>> (retrieved 24.05.10).
- Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., & Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, 59, 54–96.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59–73.
- Holcomb, P., & Neville, H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, 5, 281–312.
- Hopper, P., & Traugott, E. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.
- Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, 116, 717–751.
- Huynh, H., & Feldt, L. S. (1970). Conditions under which the mean-square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*, 65, 1582–1589.
- Iggesen, O. A. (2008a). Number of cases. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. <<http://wals.info/feature/49>> (chap. 49, retrieved 24.05.10).
- Iggesen, O. A. (2008b). Asymmetrical case marking. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. <<http://wals.info/feature/50>> (chap. 50, retrieved 24.05.10).
- Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, 83, 74–109.
- Janssen, U., Wiese, R., & Schlesewsky, M. (2005). Electrophysiological responses to violations of morphosyntactic and prosodic features in derived German nouns. *Journal of Neurolinguistics*, 19, 466–482.
- Johnson, R. Jr., (1986). A triarchic model of P300 amplitude. *Psychophysiology*, 23, 367–384.
- Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*, 14, 129–171.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–225.
- Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1–36.
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, 109, 68–74.
- Kornfilt, J. (1997). *Turkish*. London, NY: Routledge.
- Kretzschmar, F. (2010). The electrophysiological reality of parafoveal processing: On the validity of language-related ERPs in natural reading. Unpublished doctoral dissertation, University of Marburg.
- Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*, 21, 489–530.
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentence. evidence from event-related potentials. *Brain and Language*, 100, 223–237.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117–129.
- Kutas, M., & Federmeier, K. (2000). 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 incongruity. *Science*, 207, 203–205.
- Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology*, 108, 456–471.
- Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 659–724). London: Elsevier.
- Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933.
- Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457–489). New York: Academic Press.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley/Los Angeles, CA: The University of California Press.
- Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, 32, 169–192.
- Luke, K.-K., Liu, H.-L., Wai, Y.-Y., Wan, Y.-L., & Tan, L. H. (2002). Functional anatomy of syntactic and semantic processing in language comprehension. *Human Brain Mapping*, 16, 133–145.
- MacWhinney, B., & Bates, E. (Eds.) (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
- MacWhinney, B., Bates, E., & Kiegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127–150.
- Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Research*, 1096, 163–172.
- McDonald, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology*, 41, 317–335.
- Miyamoto, E. T. (2002). Case markers as clause boundary inducers in Japanese. *Journal of Psycholinguistic Research*, 31, 307–347.
- Nakagome, K., Takazawa, S., Kanno, O., Hagiwara, H., Nakajima, H., Ito, K., et al. (2001). A topographical study of ERP correlates of semantic and syntactic violations in the Japanese language. *Psychophysiology*, 38, 304–315.
- 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., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, 8, 507–526.
- Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond* (pp. 271–308). Hove: Psychology Press.
- Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W., & Schlesewsky, M. (2008). The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain and Language*, 105, 112–133.
- Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, 89, 506–540.
- Roehm, D., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2007a). The internal structure of the N400: frequency characteristics of a language-related ERP component. *Chaos and Complexity Letters*, 2, 365–395.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007b). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19, 1259–1274.
- Rögnvaldsson, E., & Thráinsson, H. (1990). On Icelandic word order once more. In J. Maling (Ed.), *Modern Icelandic syntax*. San Diego: Academic Press.
- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2009). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn.2009.21370.

- Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2009). When semantic P600s turn into N400s: On cross-linguistic differences in online verb-argument linking. In M. Horne, M. Lindgren, M. Roll, K. Alter, & J. v. K. Torkildsen (Eds.), *Brain Talk: Discourse with and in the brain. Papers from the first Birgit Rausing Language Program Conference in Linguistics* (pp. 75–97). Lund: Birgit Rausing Language Program.
- Schumacher, P. B. (in press). The hepatitis called.: Electrophysiological evidence for enriched composition. In J. Meibauer & M. Steinbach (Eds.), *Experimental Pragmatics/semantics*. Amsterdam: John Benjamins.
- Steele, S. (1978). Word order variations: A typological study. In J. H. Greenberg (Ed.), *Universals of human language, Vol. 4: Syntax* (pp. 585–623). Stanford, CA: Stanford University Press.
- Stroud, C. (2008). *Structural and semantic selectivity in the electrophysiology of sentence comprehension*. Unpublished doctoral dissertation, University of Maryland.
- Sun, C.-F., & Givón, T. (1985). On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language*, 61, 329–351.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.
- Traxler, M., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90.
- Traxler, M., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204–224.
- van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18, 1181–1197.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241–255.
- Van Valin, R. D. Jr. (1991). Another look at Icelandic case marking and grammatical relations. *Natural Language and Linguistic Theory*, 9, 145–194.
- Van Valin, R. D. Jr., (2005). *Exploring the syntax- semantics interface*. Cambridge: Cambridge University Press.
- Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of the P3. *Behavioral and Brain Sciences*, 11, 343–427.
- Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, 34, 131–156.
- Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150–163.
- Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75, 8–18.
- Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the 'subject'-preference. Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes*, 24, 1180–1226.
- Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (in press). The role of animacy in online argument interpretation in Chinese. In P. de Swart & M. Lamers (Eds.), *Case, word order, and prominence. Psycholinguistic and theoretical approaches to argument structure*. Berlin: Springer.
- Wolff, S., Schlesewsky, M., Hirotani, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, 107, 133–157.
- Ye, Z., Luo, Y.-j., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research*, 1071, 186–196.
- Ye, Z., Zhan, W., & Zhou, X. (2007). The semantic processing of syntactic structure in sentence comprehension: An ERP study. *Brain Research*, 1142, 135–145.