Research Report

Finding needles in haystacks: Symbolic resonance analysis of event-related potentials unveils different processing demands

Stefan Frisch[a,b,*], Peter beim Graben[b,c]

[a]Max-Planck-Institute of Human Cognitive and Brain Sciences, Leipzig, Germany
[b]Institute of Linguistics, University of Potsdam, Germany
[c]Institute of Physics, University of Potsdam, Germany

Abstract

Previous ERP studies have found an N400–P600 pattern in sentences in which the number of arguments does not match the number of arguments that the verb can take. In the present study, we elaborate on this question by investigating whether the case of the mismatching object argument in German (accusative/direct object versus dative/indirect object) affects processing differently. In general, both types of mismatches elicited a biphasic N400–P600 response in the ERP. However, traditional voltage average analysis was unable to reveal differences between the two mismatching conditions, that is, between a mismatching accusative versus dative. Therefore, we employed a recently developed method on ERP data analysis, the *symbolic resonance analysis* (SRA), where EEG epochs are symbolically encoded in sequences of three symbols depending on a given parameter, the encoding threshold. We found a larger proportion of threshold crossing events with negative polarity in the N400 time window for a mismatching dative argument compared to a mismatching accusative argument. By contrast, the proportion of threshold crossing events with positive polarity was smaller for dative in the P600 time window. We argue that this difference is due to the phenomenon of "free dative" in German. This result also shows that the SRA provides a useful tool for revealing ERP differences that cannot be discovered using the traditional voltage average analysis.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Much has been written on the many different types of linguistic information that have to be brought together when we understand a sentence. Generally, it is undisputed that the verb and its arguments constitute the "core" of a sentence [32]. Thus, mapping the arguments in a sentence to the verb slots is the crucial point in deriving a coherent sentential interpretation, that is, for answering the question of *who is doing what to whom* as for example in a sentence such as (1).

(1). Peter visited Mary at night.

We immediately understand that the sentence tells us something about a visiting event and that this event has two participants, namely Peter and Mary. Moreover, we also do not hesitate to conclude that Peter is doing the visiting (i.e., that he is the *agent*), whereas Mary is the one being visited (i.e., that she is the *patient*). In other words, we would not doubt that it is Peter who visits Mary and not vice versa. But our feeling that this interpretation is derived fast and effortlessly should not obscure the fact that much information processing has to be performed by our brains in order to achieve it. Accordingly, this central issue has led to many studies in both psycholinguistics (see [15], for an overview) as well as aphasiology (see [17], for an overview). In order

---

* Corresponding author. Max-Planck Institute of Human Cognitive and Brain Sciences, P.O. Box 500 355, D-04303 Leipzig, Germany. Fax: +49 341 99 40 113.
E-mail address: frisch@cbs.mpg.de (S. Frisch).

to characterize this type of information more closely, we will consider a sentence such as (2), which can immediately be judged to be unacceptable in English.

(2). *Peter snored Mary at night.

Since the surface structure of (2) is identical to (1), the reason for the unacceptability must lie in different properties of the verb *to snore* in (2) compared to the verb *to visit* in (1). More specifically, a verb such as *to visit* can take a subject argument (*Peter*) and an object argument (*Mary*), whereas an object is not possible in a verb such as *to snore*, but only a subject (*Peter*). Interestingly, this difference can be described not only syntactically (i.e., with respect to sentence structure), but also in terms of semantics (i.e., with respect to meaning/interpretation), in that the event expressed by the verb in (1) has two participants (one doing the visiting and one being visited), whereas the snoring event in (2) has only one participant (one doing the snoring). Obviously, our language processing routines have to determine whether the number of arguments in a sentence matches with the number of arguments a verb can take, and a mismatch (such as in (2)) induces both syntactic as well as semantic problems in processing. It is therefore not unexpected that electrophysiological patterns in response to a mismatch such as in (2) seem to reflect this twofold nature of the violation. Event-related brain potential (ERP) studies in both English [29] and German [11,15] sentences in which an argument could not be integrated into the verb's argument structure have been reported to elicit a biphasic N400–P600 response. Friederici and Frisch [11] have presented sentences with a subject (SUB) and a direct object (OBJ), but with an intransitive verb (i.e., only allowing for a subject, but no object) such as in (3).

(3). *Paul weiß, dass der Chemiker den Physiker emigrierte...
Paul knows that [the chemist]SUB [the physicist]OBJ emigrated.

On the mismatching verb, they found an N400 followed by a P600 compared to a condition with a transitive verb. Similar to Osterhout et al. [29], they interpreted the N400 for these types of violation as an indicator of the semantic anomaly induced by a thematic role mismatch, i.e., by the fact that there are more arguments that need a thematic role than there are roles provided by the verb. The P600 was seen to indicate the fact that a transitive structure is built up which is however not licensed by the lexicon information of the (intransitive) verb (cf. [10,15]).

A numerical correspondence between the arguments in a sentence and the argument structure information specified in the verb's lexical entry, however, does not suffice to determine *who is doing what to whom*. What is necessary in addition is that the different thematic roles of the verb are correctly assigned to the different arguments in the sentence. In a sentence such as (1), it is important to know that the verb provides two roles (*agent* and *patient*) and that there are two candidates to receive those roles (*Peter* and *Mary*),

but this alone does not tell us that Peter is doing the visiting and Mary is being visited and not vice versa. The arguments have to be systematically assigned (*mapped*) to the respective syntactic constituents. Thereby, the respective thematic "hierarchization" of the arguments has to be preserved, that is, the fact that verb arguments are distinguishable on thematic dimensions such as *control* [30] (with the agent *Peter* in (1) having more control over the visiting event than the patient *Mary*). There are important cues in the linguistic input that allow us to determine the correct assignment effortlessly, at least in most cases. These cues are not the same for all languages, and German and English provide good examples for languages with differing cues. In an English (declarative) sentence such as (1), the linear order of the arguments is crucial: *Peter* is (linearly, in a left-to-right sequence) the first argument in the clause and therefore the higher argument (with more agentive properties) compared to the second argument *Mary* that is thematically lower. In German, by contrast, case morphology is the crucial cue for hierarchizing arguments (cf. [7,13]), as it is in many other languages with a relatively free word order. In a sentence such as (3), the subject indeed precedes the object, but in (4), the first argument bears accusative (ACC) case (direct object), whereas the second argument is marked with the subject case nominative (NOM).

(4). Den Dichter hat der Arzt zuerst besucht.
  [the poet]ACC has [the doctor]NOM first visited.

Assigning the thematic roles in (4) on the basis of linear order would lead to a completely different (and wrong) interpretation of the sentence. That case information of arguments in German is used immediately for syntactic analysis as well as thematic interpretation (i.e., hierarchization) can also be traced neurophysiologically [7,13,14]. For example, it can be shown that constructions with two identically case marked arguments induce a biphasic pattern of an N400 (reflecting thematic interpretation problems) and a subsequent P600 (reflecting ill-formedness) compared to sentences in which the arguments can be hierarchized on the basis of different case markings [13,14].

With respect to ditransitive constructions, that is, sentences with two object arguments, English also strictly relies on argument order. In (5a), it is clear that the tickets are given and the brother must be the recipient of the tickets. Swapping the argument noun phrases (NPs) would result in a semantic anomaly, see (5b).

(5a). Mary gave her brother the tickets.

(5b). ??Mary gave the tickets her brother.

In German, by contrast, this is again different, as it is again the case marking of the arguments (rather than their linear order) which determines their grammatical function: A dative (DAT) marked NP is the indirect object, whereas an accusative NP is the direct object. This is independent of

the linear sequence of the two since (6a) and (6b) have the same meaning.

(6a). Der Junge hat dem Freund das Buch gegeben.
   [the boy]NOM has [the friend]DAT [the book]ACC given.
   'the boy has given the book to his friend'.

(6b). Der Junge hat das Buch dem Freund gegeben.
   [the boy]NOM has [the book]ACC [the friend]DAT given.
   'the boy has given the book to his friend'.

With respect to the earlier studies testing constructions with a surplus argument [11,15], there is – at first sight – no reason not to expect a biphasic N400–P600 effect for a transitive verb that occurs in a sentence with three arguments such as (7).

(7). *Der Junge hat dem Freund das Buch gekannt.
   [the boy]NOM has [the friend]DAT [the book]ACC known.

However, although the argument "dem Freund" in (7) cannot be integrated into the argument structure of a verb such as *kennen* (*to know*), the question arises whether the resulting processing problems might differ from those induced by the verb in (3). Such an assumption could be derived from the fact that the mismatching NP bears dative (indirect object) case and that dative case in German can be used in order to expand the argument structure of a transitive (nominative–accusative) verb, see (8).

(8). Anna hat ihrem Bruder das Motorrad repariert.
   Anna has [her brother]DAT [the motorcycle]ACC repaired.
   'Anna has repaired the motorcycle for her brother'.

The verb *reparieren* (*to repair*) itself has only two arguments (a repairer and something being repaired), according to the criterion of semantic necessity [33]. Adding a so-called *benefactive dative* serves to indicate someone who profits from (or is harmed by) the event expressed by the verb. Adding an argument expressing a beneficent is possible in many languages, although it is achieved by different syntactic means. In English, for example, such a beneficent is not realized as an indirect object, but by means of a prepositional phrase (PP) with "for" (see (8)). In German, by contrast, it is realized as an indirect object with dative case, which behaves syntactically exactly like a "true" (i.e., semantically necessary, cf. [33]) dative argument of a ditransitive verb such as *dem Freund* in (6a) and (6b) and is therefore syntactically indistinguishable from the latter.[1] There is some controversy concerning the exact conditions which have to be met so that a benefactive dative in German can be added (such as in (8)) or whether such a procedure results in an unacceptable construction (such as in (7)) [34,37]. It has been argued, for example, that a benefactive dative is not possible if such a status of the dative argument

was not intended by the agent [37], as it is normally not the case with experiencer verbs. Furthermore, it was proposed that (8) is not possible when the benefactive dative is not physically affected [34]. Although none of these accounts gives an exhaustive characterization, they converge in the view that the benefactive dative has to fulfill restrictions by the verb that are by and large semantic in nature (cf. [34,36,37]). By contrast, there is no such possibility of adding an accusative object in German sentences with transitive verbs that mark their object irregularly with dative case, such as in (9).

(9). *Anna hat ihrem Bruder die Hausaufgaben geholfen.
   Anna has [her brother]DAT [the homework]ACC helped.

This principal possibility – however limited – for adding a dative–but not an accusative–argument to a transitive relation might affect the ERP patterns for an extra-dative compared to an extra-accusative differently. With respect to the present study, therefore, the following questions arise: First, is the biphasic N400–P600 pattern as found for argument structure violations in intransitive (Frisch et al., 2004) and transitive structures (Friederici and Frisch, 2000) also found in ditransitives such as (7) or (9)? Second, do the linguistic differences between adding a dative and adding an accusative affect the N400–P600 differently for an incorrect dative compared to an incorrect accusative?

2. The present study

From the fact that adding a dative argument in German is only semantically restricted whereas adding an accusative is both syntactically and semantically impossible, one might expect different electrophysiological responses for sentences with a mismatching dative (inducing mainly semantic processing problems) compared to sentences in which an accusative marked argument cannot be integrated (inducing both semantic and syntactic processing problems). Although sentences with a mismatching argument have been consistently found to elicit biphasic N400–P600 patterns in ERPs [11,15], these patterns might be different depending on the type of argument, that is, if the mismatching argument bears accusative or dative case. Therefore, we presented sentences with a surplus accusative as well as with a surplus dative object and compared the ERP responses on the verb to those

---

[1] It has been argued that, in German, there is a similar way to add a beneficent by means of a PP as in English and that this PP is equivalent to a benefactive dative [19]. This was taken as evidence that benefactive dative in German is an adjunct rather than an argument. However, it can be shown that *Anna hat das Motorrad für ihren Bruder repariert* (*Anna has repaired the motorcycle for her brother*) is not a paraphrase of Sentence 8 [9]. Furthermore, the fact that benefactives are affected by all syntactic restrictions (for example, *recipient passive*, *topicalization*, etc.) in exactly the same way as "real" argument datives can be taken as a criterion to give them *argument status* [9].

elicited by correct sentences. Seeing that traditional voltage average analysis was only able to replicate the generally expected ERP pattern, but not the fine-grained differences between the two mismatching conditions, a newly developed analysis procedure, the symbolic resonance analysis [4], was employed.

3. Materials and methods

3.1. Subjects

Sixteen monolingual native speakers of German (mean age 23 years, six female), all students from the University of Leipzig, participated in the experiment. They were all right-handed according to Oldfield [27] and had normal or corrected-to-normal vision. They were naïve with respect to the aims of the study and were paid for their participation.

3.2. Materials

All sentences were created out of 40 sets of three NPs and a sentence final PP. Each of these sets had four different verbs, two ditransitive ones that completed the sentence correctly, a transitive verb that marked its sole object in the accusative, and one whose sole object had to bear dative case. The second ditransitive verb was used to create another 80 sentences in order to balance the number of correct and incorrect conditions. Sentences in this latter condition were treated as fillers and were not analyzed any further as the verbs in this condition were not matched with respect to lexical parameters. By contrast, such a matching was done for the three groups of critical verbs that were kept similar in logarithmic lemma frequency as determined on the basis of the CELEX database (cf. [1]). The mean for correct verbs was 1.19, for the verbs not subcategorizing a dative 1.28, and for the verbs that could not take an accusative 1.10. An ANOVA performed for these frequencies revealed an $F < 1$ and no differences in direct single comparisons [correct vs. dative: $F < 1$; correct vs. accusative: $F < 1$; dative vs. accusative: $F(1,78) = 1.62$, $P = 0.21$].

Seeing that German declarative sentences have an unmarked order with respect to the object arguments (with the dative object preceding the accusative object, see (5a) and (5b)), we were faced with the possible confound that the distance between the mismatching argument and the critical verb was different for the dative and the accusative object. We therefore decided to use welcher/which-questions where the which-constituent is always in the same position (i.e., at the beginning of the subclause) irrespective of its case. Furthermore, we varied the argument in the which-position systematically between direct/accusative and indirect/dative object, as exemplified in (10a) and (10b).

(10). Jochen weiß,
    Jochen knows

(a) [welchen Betrag]ACC [der Bläser]NOM [dem Geiger]DAT neulich VERB-te
    [which amount]ACC [the trumpeter]NOM [the violinist] DAT recently VERB-ed
(b) [welchem Geiger]DAT [der Bläser]NOM [den Betrag]ACC neulich VERB-te
    [which violinist]DAT [the trumpeter]NOM [the amount] ACC recently VERB-ed
    bei jener Reise nach Paris.
    during that visit to Paris.

Examples for the VERB-position for each of the three critical conditions are provided in (11), (12), and (13):

(11). *COR: correct (ditransitive) verb:* borgte (*lent*).

(12). *DAT: verb which cannot take a dative argument:* verbrauchte (*consumed*).

(13). *ACC: verb which cannot take an accusative argument:* half (*helped*).

After the verb, a complex (and therefore preferably extraposed) prepositional phrase (PP) such as "during that visit to Paris" in (10) was added in order to avoid confounding sentence final wrap-up effects (cf. [11,28]).

3.3. Procedure

There were 80 sentences in each of the three critical conditions (40 with the dative argument in which-position and 40 with the accusative argument in which-position). All sentences were presented in randomized order in the center of a computer monitor as words or phrases, respectively. All three NPs and the second PP were presented for 500 ms, the first PP for 550 ms. All other items were presented word-by-word for 400 ms. The ISI was 100 ms. 800 ms after the sentence final PP, the subjects were asked to judge the acceptability of the sentence within a 2500-ms interval by pressing a button. 1000 ms after their response, the next trial began.

The EEG was recorded by means of 59 Ag/AgCl electrodes with a sampling rate of 250 Hz and was referenced to the left mastoid (re-referenced to linked mastoids off-line). In order to control for eye movement artifacts, a horizontal electro-oculogram (EOG) was monitored from electrodes at the outer canthus of each eye and a vertical EOG from two electrodes located above and below the subject's right eye. Electrode impedances were kept below 5 kΩ. EEG and EOG channels were recorded continuously with a band pass filter from DC to 30 Hz with a digitization rate of 250 Hz.

3.4. EEG data analysis

EEG data were high pass filtered with a cutoff frequency of 0.4 Hz in order to compensate for drifts. Only correctly performed trials without ocular or amplifier saturation

artifacts entered the EEG data analysis (83% of all trials, evenly distributed across conditions).

3.4.1. Voltage average data

ERPs were averaged in a 1300-ms time window (relative to the onset of the critical verb) and aligned to a 200-ms pre-stimulus baseline. ERPs were additionally filtered with a 10-Hz low pass filter for presentation purposes only.

3.4.2. Symbolic resonance analysis

3.4.2.1. General idea. The naming conventions of ERP research suggest that only polarity and latency of a voltage deflection are to be taken into consideration even at the level of single EEG trials. This has, of course, some tradition in ERP analysis: by counting the number of EEG epochs which have a positive or negative voltage value at a certain instance of time, one obtains the *polarity histogram* reflecting the intertrial coherence of the ERP [8]. Lehmann [22] suggested to consider only positive and negative maximal field values. The common overlap of their time ranges across trials yields again a measure of the intertrial coherence. However, these coarse-graining techniques were lacking a theoretical foundation that has recently been provided by beim Graben et al. [6] (cf. [3,5,16]) in the framework of symbolic dynamics [18,23]. In this branch of nonlinear science, measured or predicted time series of dynamical systems are mapped onto sequences of very few symbols by partitioning the range of values that is assumed by the data. In the case of ERP data, for example, one can take the pre-stimulus baseline as a threshold and assign the symbols "+" or "−" to each sample point if the measured EEG in a single trial is above or below the baseline at this time point, respectively, thus obtaining a sequence of "+" and "−".[2] After collecting all the sequences corresponding to one experimental condition in an ensemble, the number of "+" or "−" symbols across all trials yields the polarity histogram [8]. beim Graben et al. [6] have shown that these polarity histograms can be formally captured by probability measures of *cylinder sets*, which are subsets drawn from an ensemble of sequences having a common building block, which is called a *word*. beim Graben et al. [6] have further argued that ERP components are characterized by large cylinder sets corresponding to a particular word. The likelihood of these cylinders is then assessed by generalized polarity histograms, the *word statistics* [5,16], and by information theoretic measures such as the Shannon entropy or measures of complexity [2,31], which serve as indicators of the intertrial coherence [24].

In particular, Lehmann's idea to encode the maxima and minima of the EEG time series has led us to developing the

symbolic resonance analysis (SRA) of ERP data [4]. In the approach pursued in the present paper, three symbols instead of two are assigned to the EEG data at each sample point by introducing two thresholds which partition the range of voltage values of the EEG into three intervals. This results in the following encoding rule (cf. [4]):

(14). One symbol ("0") is assigned to all sample points below the lower threshold, the second one ("1") for the sample points above the lower and below the upper threshold, and the third ("2") for all data points above the upper threshold.

The distance between both thresholds can be varied, but the absolute value, abbreviated θ, of both thresholds must be the same (i.e., both thresholds are equidistant from the baseline and only different in their signs).[3]

An EEG epoch is thereby mapped onto a sequence of "0"s, "1"s, and "2"s. For instance, "10122" means that the signal was between the thresholds at times $t = 1$ and $t = 3$, below the lower threshold at $t = 2$, and above the upper threshold at $t = 4$ and $t = 5$. Since time is represented by discrete sampling points, for a sample rate of, e.g., 250 Hz, $t = 1$ corresponds to 4 ms, $t = 2$ to 8 ms, and so on. It is clear that the symbolic representation of the EEG epoch depends on the chosen threshold. If θ is too small, the intermediary symbol "1" will not occur often since the signal is oscillating between its maxima and minima. If, on the other hand, θ is too large, that is, larger than the maximum of the absolute value of the upper and lower boundaries of the signal, the symbols "0" and "2" will never occur, and one observes only sequences of "1"s.

Let us assume that the absolute value θ of the threshold is slightly larger than the amplitude of the ERP content of the signal, which results from the ensemble average of the baseline aligned EEG epochs. Employing the encoding rule (14) at the signal yields a sequence of only "1"s since the thresholds $-\theta$ and $+\theta$ will not be exceeded by the signal at all.

However, because a single epoch of the raw EEG is regarded in the traditional voltage averaging analysis as the sum of the underlying nonstationary ERP wave and some noise, rule (14) applied to the raw EEG yields "2"s whenever the noise drives the signal across the upper threshold $+\theta$ near the maxima of the ERP, and "0"s if the

---

[2] Note that the particular symbols assigned to the measurements are completely arbitrary. So one might also use "0" instead of "−" and "1" instead of "+", or "a" instead of "−" and "b" instead of "+", or any other representation.

[3] As one of the referees suggested, one could compute the standard deviation of the signal in the pre-stimulus time window and express the threshold in standard deviation units. We actually attempted this during the development of the SRA. Compared to the use of absolute voltages, this approach has one serious disadvantage: the standard deviation is obtained for each single EEG time series, and the standard deviations of different trials could thus differ significantly. As a consequence, using standard deviation units, one would abandon any information about the amplitude of the ERP. As we will point out below, the critical encoding threshold represents this important information. Thus, absolute voltage thresholds make the SRA results more compatible with the averaged ERPs than standard deviation thresholds would do.

sum of the ERP and the noise is smaller than the lower threshold $-\theta$, which is the case around the minima of the ERP. Thus, the local maxima and minima of the ERP are encoded by "2"s and "0"s, respectively. This is illustrated in Fig. 1, where we display in Fig. 1(A) a nonstationary ERP-like test signal (a Bessel function) having a local maximum at $t = 0$ with amplitude 1 and two local minima around $t = -4$ and $t = +4$. Adding noise to this subthreshold signal (Fig. 1(B)) makes threshold crossing events probable, resulting in the symbol "2" around $t = 0$ and the symbol "0" around $t = -4$ and $t = +4$.

As we have already mentioned, while the background EEG superimposed with the ERP is commonly regarded as being detrimental noise in the customary ERP analysis that has to be eliminated by averaging, our three-symbol encoding rule (14) utilizes it constructively to drive the subthreshold ERP signal across the encoding thresholds.

By encoding each EEG epoch for a fixed threshold, we obtain a set of sequences, which can be arranged as an array of the symbols "0", "1", and "2". Table 1 displays such a

Fig. 1. (A) An ERP like test signal (the zero-order Bessel function of the first kind) with a positive voltage deflection at $t = 0$. (B) The test signal superimposed with Gaussian white noise of variance 0.64 and encoding thresholds yielding a symbolic dynamics with "0", "1", and "2".

possible array, where the time is given by the columns while the rows denote the trial index.

The next step of the SRA is determining the word statistics (i.e., the polarity histogram). We restrict ourselves to the statistics of the three symbols across trials. The word statistics count the relative frequencies of "0"s, "1"s, and "2"s depending on time. Considering Table 1 again, we observe that at $t = 1$, the frequency of "0"s is $P_0 = 0.25$, the frequency of "1"s is $P_1 = 0.50$ and the frequency of "2"s is $P_2 = 0.25$. At $t = 2$, we have $P_0 = 0.50$, $P_1 = 0$, $P_2 = 0.50$, and so on. Table 2 shows the resulting symbol distributions.

Let us assume that there is a local maximum in the ERP at $t = 4$ with no effects at other times for this example. beim Graben and Kurths [4] presented an algorithm to transform the distribution of three symbols to a distribution of only two symbols, "0" and "2"s, where "0" denotes a local minimum in the ERP, while "2" denotes a local maximum.[4] This transformation, which is inspired by the theory of spin lattice models in statistical mechanics [35] works in the following way: The differences of the symbol frequencies $M_0 = P_0 - P_1$ and $M_2 = P_2 - P_1$ are regarded as competing "magnetic fields" (so-called mean-fields) which act at the symbol distribution across the columns of Table 1 trying to flip the "undecided" symbol "1" either into a "0", if $M_0 \geq 0 > M_2$ (i.e., there are more "0"s than "2"s, which thereby win the competition), or into a "2", if $M_0 < 0 \leq M_2$ (when there are more "2"s than "0"s). We present the magnetic field strengths for the example from Table 1 in Table 3.

By looking at the mean-fields in Table 3, we can now decide how the "1"s of Table 1 must be flipped into "0"s or "2"s, according to the spin-flip transformation. At $t = 1$ we have $M_2 \leq M_0 < 0$, where the transform is still undefined. We close this gap by saying that, if $M_2 \leq M_0 < 0$ or $M_2 \geq M_0 > 0$, the "1"s should be equally converted into the same number of "0"s and "2"s, i.e., one half of the "1"s becomes "0"s while the other half becomes "2"s. Thus, at $t = 1$, we obtain column one of Table 4 where the "1" in the first row of Table 1 is converted into "0" and the "1" in row 3 is flipped into a "2". The particular symbol that has to be flipped was chosen randomly. At $t = 2$ we see that $M_2 \geq M_0 > 0$, the second case where the "1"s are equally converted into "0"s and "2"s. However, in our example, there is no "1" to which the transformation applies so that nothing happens. At $t = 3$ and also at $t = 5$, we encounter the same situation as at $t = 1$. However, at $t = 4$, there is a real winner of the competition, namely, "2", where $M_0 < 0 \leq M_2$. Here, all "1"s are flipped into "2"s. The mean-field transformation might remind the reader of the game "Reversi", where single-colored chips have to be flipped when they are trapped by chips of the converse color. We will therefore call this transformation the *Reversi transformation* (see Table 4).
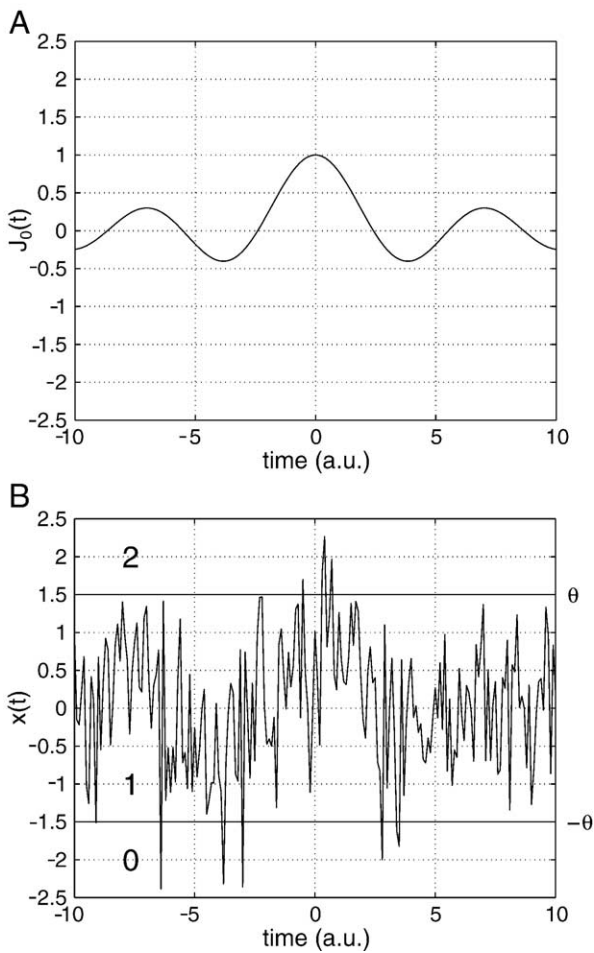
---

[4] Our presentation of the mean-field transformation here deviates a bit from that given by beim Graben and Kurths [4] in order to make the idea more clear.

Table 1
Example of a three-symbol ERP dynamics

Epoch\t	1	2	3	4	5
1	1	0	1	2	2
2	0	2	2	1	1
3	1	0	0	2	0
4	2	2	1	2	1

Rows denote EEG trials, columns denote sample points.

After employing the *Reversi transformation*, we obtain an array of only "0"s and "2"s (cf. [4]) where "0" denotes "negative polarity" and "2" denotes "positive polarity" with respect to the chosen threshold. From this recoded array, we determine again the word statistics that are shown in Table 5 for the present example.

The symbols "0" and "2" are obviously uniformly distributed for $t = 1, 2, 3, 5$ but not at $t = 4$, where the distribution is highly degenerated. As we have pointed out above, in our example model, an ERP deflection was assumed at $t = 4$, thus the spin-flip transformation enhances threshold crossing events in the positive and negative direction in a strongly nonlinear way, while suppressing random fluctuations around the baseline which leads either to a large number of "1"s in the original three-symbol encoding for large thresholds, or to an almost uniform distribution of "0"s and "2"s for small thresholds. Both cases are mapped by the *Reversi transformation* onto a uniform distribution of "0"s and "2".

From the word statistics, we then derive the running cylinder entropies [6] that are presented in Table 6 for the example data.

Entropy is a measure of uncertainty of a given probability distribution. It reaches its maximum value 1.0 for uniformly distributed events and it assumes its minimum 0.0 if there is only one certain event with probability 1.0 (cf. [31]). Entropies of symbol distributions measure the amount of order in the system at one instance of time and thereby its intertrial coherence. Since ERP components are reflected by highly degenerated word statistics, the corresponding cylinder entropies generally decrease within the time range of an ERP. It was therefore tempting to relate the amplitude of an entropy drop to the signal-to-noise ratio (SNR) of the EEG data. This has been accomplished by beim Graben [3], who found that the time-averaged entropy is inversely proportional to the SNR incremented by a constant. Applied to our example from Table 6, we state that the SNR of the signal is zero at $t = 1, 2, 3, 5$ and high at $t = 4$.

So far, we have described the SRA for a fixed encoding threshold that is approximately as large as the amplitude of a

Table 3
Magnetic mean-fields obtained from the word statistics of Table 2

Prob\t	1	2	3	4	5
M_0	−0.25	0.50	−0.25	−0.25	−0.25
M_2	−0.25	0.50	−0.25	0.50	−0.25

voltage ERP. But how can the appropriate threshold be found? The answer is, by trial-and-error. We therefore employ the abovementioned algorithm over a range of reasonably chosen threshold parameters. After computing the SNR within some interesting time window, we plot these results against the encoding threshold. We have seen that if the threshold is too small, we observe mainly "0"s and "2"s in the original three-symbol encoding, which are mapped onto a uniform distribution of "0"s and "2"s by the *Reversi transformation*, as was the case at $t = 2$ in our example. This distribution has maximal entropy and therefore a low SNR. On the other hand, if the threshold is too large, prohibiting any threshold crossing events, we observe almost all "1"s in the three-symbol encoding, which are again mapped onto a uniform distribution of "0"s and "2"s by the *Reversi transformation*. This was the case at $t = 1$, $t = 3$, and $t = 5$ in our example. As a result, we have a low SNR here as well. However, an ERP component gives rise to a maximum of the SNR at a critical threshold $\theta*$, indicating most probable threshold crossing events in the positive or negative direction, or, technically speaking, an *aperiodic stochastic resonance effect* [4,25,26].

The critical threshold, $\theta*$, where this resonance takes place, depends, of course, on the time window where the SNR is computed. This time window can be heuristically determined by the latency of the averaged voltage ERP. Slight variations of the onset and the end of the window yield only slight differences between the resonance curves; the dependency is therefore rather robust against arbitrary changes of the analysis parameters. Moreover, one has to observe that the determination of $\theta*$ compensates for the alleged loss of information entailed by the symbolic coarse-graining. The critical threshold contains all information about the "true" amplitude of the ERP. In contrast to the voltage-averaged ERP, where amplitude and coherence information are mixed up into one dimension, the symbolic resonance analysis pulls both kinds of information apart into two different dimensions: amplitude is represented by the critical encoding threshold, whereas intertrial coherence is contained in the time-dependent word statistics or cylinder entropy. The SRA is also robust against outliers in the EEG, whereas the amplitudes of the voltage-averaged ERP are

Table 2
Word statistics of the model data from Table 1

Prob\t	1	2	3	4	5
P_0	0.25	0.50	0.25	0	0.25
P_1	0.50	0	0.50	0.25	0.50
P_2	0.25	0.50	0.25	0.75	0.25

Table 4
Mean-field transformed symbolic dynamics from Table 1

Epoch\t	1	2	3	4	5
1	0	0	2	2	2
2	0	2	2	2	0
3	2	0	0	2	0
4	2	2	0	2	2

Table 5
Word statistics of the mean-field transformed example data

Prob\\t	1	2	3	4	5
P'_0	0.50	0.50	0.50	0	0.50
P'_2	0.50	0.50	0.50	1	0.50

very susceptible to them. Consider two cases: First, a possible outlier is consistent with an ERP effect, i.e., the outlier has the same polarity as the ERP. By means of the coarse-graining, the outlier will be represented by the very same symbol (either "0" or "2") as the ERP, but will contribute only one trial to the word statistics. On the other hand, in the averaging paradigm, an outlier is weighted by its numerical value in the ERP average, which might cause a large deviation. In the second case, when the outlier is not consistent with the ERP component, the ERP average is diminished by its numerical extent. Conversely, one inconsistent outlier-trial is treated by the *Reversi transformation* of the SRA either as being undecided ("1") or as a loser in the competition of the mean-fields. Thus, the coarse-graining of the SRA combined with the multiple-threshold analysis does not involve any renunciation of information, but is very robust against statistical outliers on the other hand.

3.4.2.2. Symbolic resonance analysis of the present EEG data. The EEG data were epoched in a time window beginning 200 ms before and ending 1300 ms after the onset of the critical verb. Each EEG epoch had been firstly encoded in sequences of the three symbols "0", "1", and "2" according to the encoding rule ((14)), after aligning their baselines to the time average of the 200-ms pre-stimulus interval. The encoding thresholds θ were tuned from 0.5 μV up to 9.8 μV in steps of 0.1 μV. Afterwards, the symbolic sequences of all subjects per threshold and per condition were swept up to the *grand epoch ensembles* (*GEE*, cf. Table 1) from which the relative frequencies of the symbols in each time slice have been determined (cf. Table 2). These three-symbol distributions were subjected to the *Reversi transformation*, leading to a distribution of two symbols "0" and "2", whose relative frequencies yield the transformed word statistics (cf. Table 5).

Fig. 2 illustrates the effect of the *Reversi transformation* by showing the original three-symbol distribution for the mismatching dative ERPs at electrode PZ in Fig. 2(A) for the encoding threshold $\theta = 4.8$ μV. The N400 is reflected by the largest frequency of "0"s around 400 ms after stimulus onset, while the P600 is indicated by the largest frequency of "2"s around 750 ms. By contrast, Fig. 2(B) shows the two-word statistics resulting from the *Reversi transformation*. In the 400-ms window, the between-threshold symbols "1" are completely reverted into "0"s due to the impact of the mean-

fields. Hence, the N400 corresponds to a highly degenerated word statistics of almost constant amplitude over the characteristic time window. Correspondingly, the between-threshold symbols "1" are flipped into the symbol "2" in the P600 time range. When the "1"s predominate the three-symbol distributions, such as before 400 ms or after 1000 ms, the resulting two-symbol distributions are uniform. Note that both the symbolic encoding rule ((14)) and the *Reversi transformation* act instantaneously and independently at each sampling point in time across all measured epochs. This assures that the word statistics of different ERP components are independent from one another. This might not be the case for other symbolic encoding procedures, such as the median threshold encoding (cf. [6]).
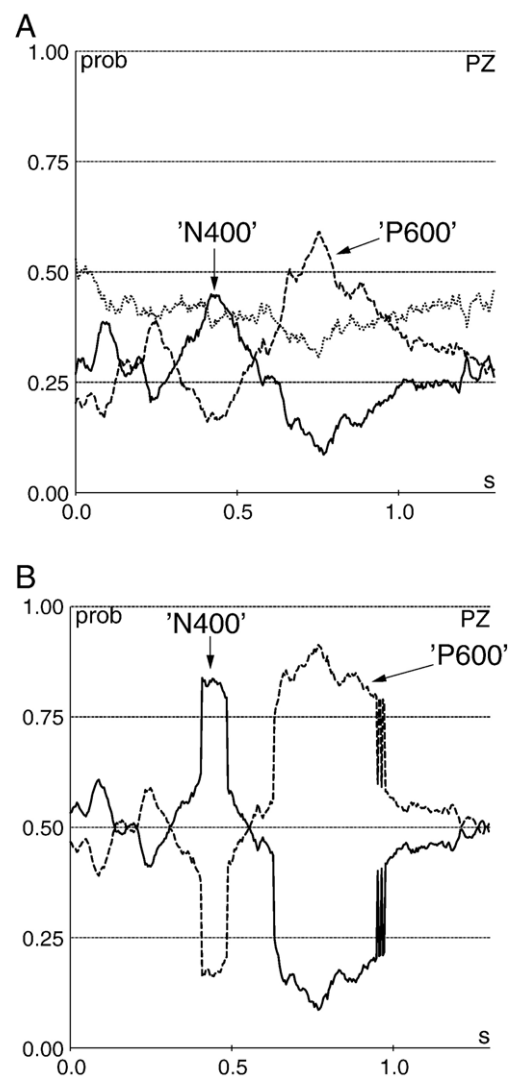

Fig. 2. (A) Relative frequencies of the symbols "0" (solid), "1" (dotted), and "2" (dashed) in the coarse of time of the grand ensemble EEG of the mismatching dative condition for encoding threshold $\theta = 4.8$ μV at electrode PZ. (B) Relative frequencies of the symbols "0" (solid), and "2" (dashed) after the spin-flipping *Reversi transformation* applied to the same data. N400 and P600 are indicated by arrows.

Table 6
Running cylinder entropies of the mean-field transformed example data

t	1	2	3	4	5
$H(t)$	1.0	1.0	1.0	0.0	1.0

In order to achieve the aim of the symbolic resonance analysis, we have to compute the estimator of the signal-to-noise ratio (SNR) [4] depending on the noise strength of the EEG. Since the noise itself is not accessible in single EEG epochs, we have systematically varied the encoding thresholds. The first step towards determining the SNR is computing the running cylinder entropies of the mean-field filtered two-symbol statistics for each condition, respectively. Fig. 3 displays the cylinder entropies of the dative condition for three different encoding thresholds: $\theta = 2.0$ μV, $\theta = 4.8$ μV, and $\theta = 6.0$ μV.

Fig. 3 illustrates the resonance effect of the three-symbol encoding combined with the *Reversi transformation*. For thresholds that are too small ($\theta = 2.0$ μV, dotted line) or too large ($\theta = 6.0$ μV, dashed line), the ERP appears either as noise or remains subthreshold, both leading to uniform distributions of the spin-flipped three-symbol encoding, thus having large entropy.

Finally, we average the cylinder entropies over the same time windows that were used for the ANOVA of the voltage-averaged ERPs, i.e., 300 to 600 ms for the N400, and 700 to 1000 ms for the P600. This yields the SNR estimators that are shown in Fig. 4.

Fig. 4 reveals particular differences in the ERP between conditions. All conditions lead to a bell-shaped resonance curve indicating the presence of a signal that is not simply noise recorded during the experiment. On contrast, both experimental conditions (accusative: dashed, dative: dotted) possess resonance curves with larger amplitude than the control condition (solid). The amplitude of the resonance indicates the intertrial coherence of the ERP, while its abscissa is a direct measure of the amplitude of the ERP signal. Fig. 4(A) shows that the ERP effect elicited by the mismatching dative has a much larger SNR and thus a larger intertrial coherence than the ERP that is related to the mismatching accusative in the N400 time window. On
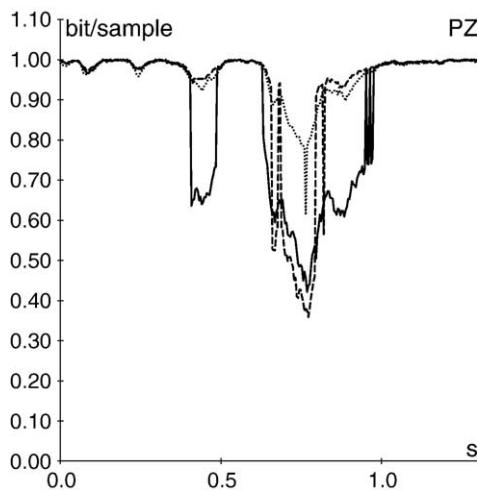

Fig. 3. Running cylinder entropies of the grand ensemble EEG of the mismatching dative condition for three different encoding thresholds $\theta = 2$ μV (dotted), $\theta = 4.8$ μV (solid), and $\theta = 6.0$ μV (dashed) at electrode PZ.
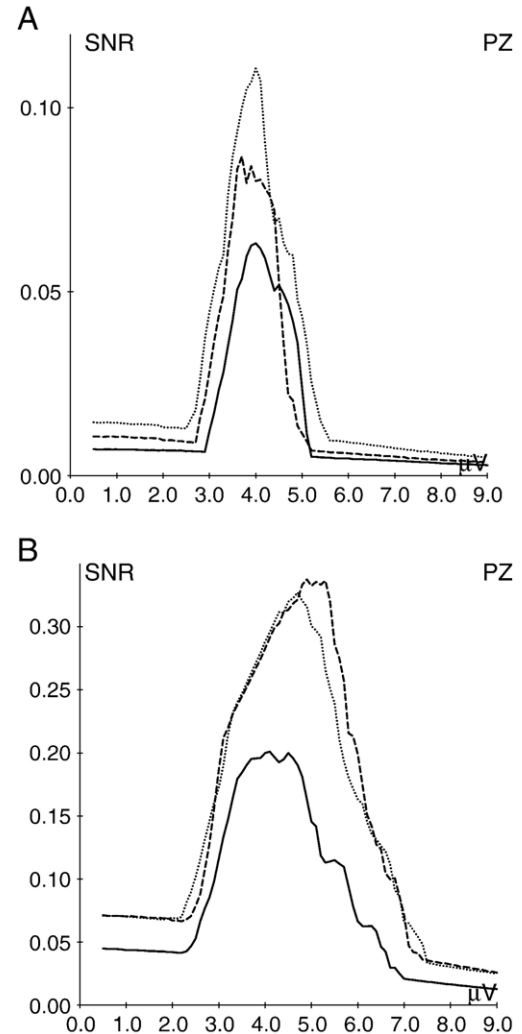

Fig. 4. (A) Resonance curves (signal-to-noise ratio against encoding thresholds) for the ERP data of three conditions obtained by averaging the cylinder entropy in the time window from 300 to 600 ms for the N400 ERP. Conditions: correct (solid), accusative (dashed), and dative (dotted). (B) Resonance curves (signal-to-noise ratio against encoding thresholds) for the ERP data of three conditions obtained by averaging the cylinder entropy in the time window from 700–1000 ms for the P600. Conditions: correct (solid), accusative (dashed), and dative (dotted).

the other hand, Fig. 4(B) discloses a contrary effect where the mismatching accusative leads to a higher SNR maximum than the mismatching dative case in the P600 time window.

Since we are interested in differences between conditions, we introduce the concept of the optimal threshold, meaning those threshold values which maximize the SNR difference of two conditions. While the optimal thresholds for the accusative condition compared to the correct condition and for the dative condition with respect to the correct condition are both around 4.0 μV in the N400 window (Fig. 4(A)), the optimal threshold for the dative condition compared to the accusative condition is just 4.8 μV, the value we have used in our illustrations. In the P600 time window, we obtain an optimal threshold of

5.3 μV at PZ between the dative condition and the accusative condition.

3.5. Statistical analysis

The statistical analyses of the voltage average data were computed in an ANOVA with repeated measures on the basis of the original data. Analyses were computed separately over the midline and the lateral electrode sites. The design consisted of a condition factor VIOLATION (VIOL) with the three levels *correct* (COR) vs. *mismatching dative* (DAT) vs. *mismatching accusative* (ACC) and a topographical factor ELECTRODE (ELEC) for the midline analysis and a factor REGION OF INTEREST (ROI) for the analysis of the lateral sites, respectively. The factor ELEC had three levels, namely, the midline electrodes *FZ* vs. *CZ* vs. *PZ*. The factor ROI had six levels, namely, the following lateral ROIs: *left-anterior* (electrodes F3, F5, FC3, FC5), *right-anterior* (electrodes F4, F6, FC4, FC6), *left-central* (electrodes C3, C5, CP3, CP5), *right-central* (electrodes C4, C6, CP4, CP6), *left-posterior* (electrodes P3, P5, PO3, PO7), and *right-posterior* (electrodes P4, P6, PO4, PO8).

When computing post hoc single comparisons between the three levels of the factor VIOLATION, the probability level was adjusted according to the modified Bonferroni procedure (cf. [21]). To protect against excessive type 1 errors, resulting from violations of sphericity, the correction proposed by Huynh and Feldt [20] was applied when evaluating effects with more than one degree of freedom in the numerator. In these cases, we report the original degrees of freedom and the corrected probability level.

The symbolic resonance data were statistically evaluated using the same ANOVA design as for the voltage average data. In order to do this, we determined the optimal thresholds for both time windows for each single electrode of the arrays displayed in Figs. 5–7. In order to allow for employing an ANOVA statistics for the symbolic resonance analysis as well, we decomposed the *Reversi transformed* GEE of two symbols into the single subject ensembles. Then, we could compute the relative frequencies of "0"s and
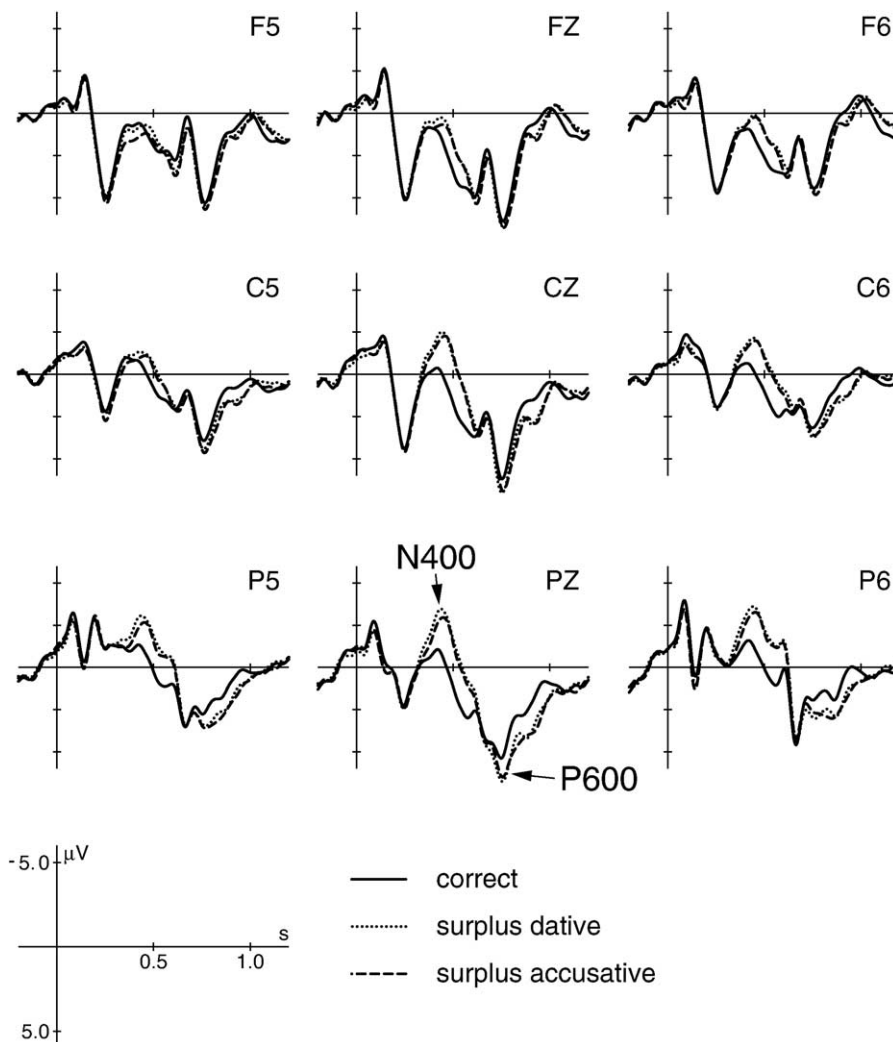

Fig. 5. Averaged ERPs from the onset of the critical item (verb, onset marked by vertical line at 0 ms) up to 1200 ms thereafter for all 16 subjects at a subset of nine electrodes. Negativity is plotted upwards.
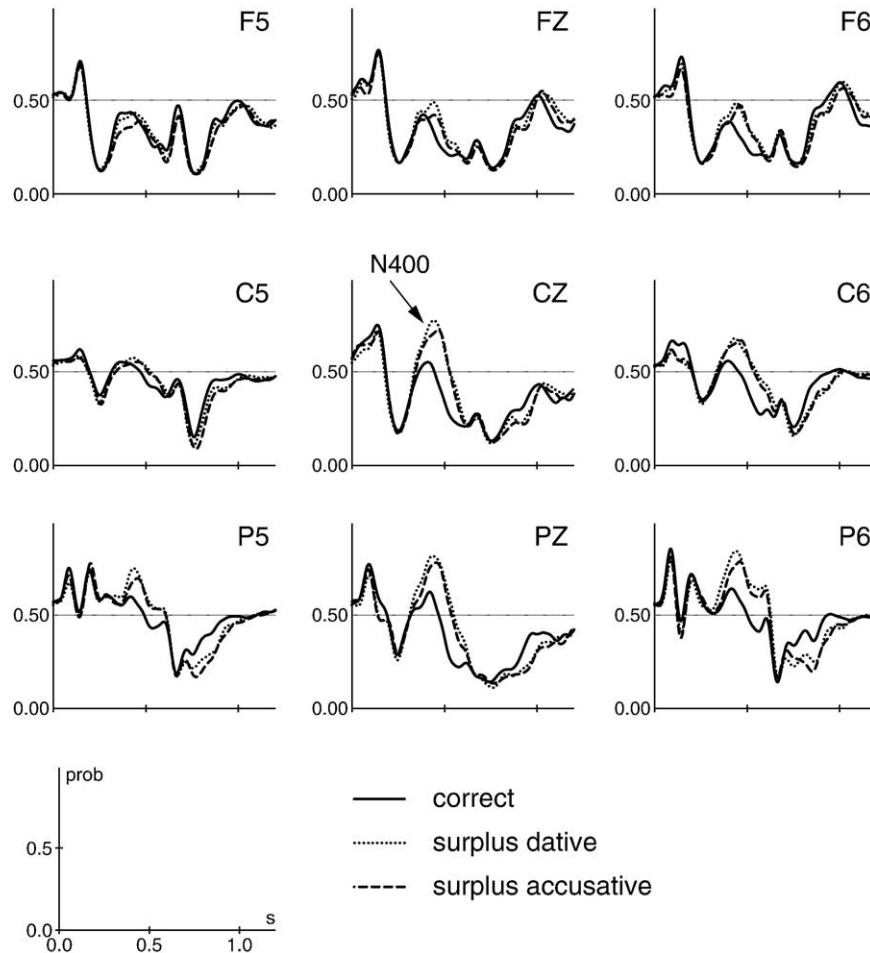
Fig. 6. Distribution of "0"s (relative frequency of trials with negative polarity) in each of the three critical conditions averaged over 16 subjects and over threshold range (I) (4.0 to 4.8 μV) from the onset of the critical item (verb, onset marked by vertical line) up to 1200 ms thereafter.

"2"s for each subject separately and computed ANOVAs with exactly the same design as for the voltage-averaged ERPs (see above). In order to reduce the impact of singular resonance events and also of the selected reference electrodes for which the optimal thresholds have been computed, we averaged the symbol statistics over a threshold window ranging from the smallest optimal threshold determined for one reference channel to the largest optimal threshold determined for another reference.

4. Results

4.1. Behavioral data

Although the acceptability judgments served only as a control task in order to make sure that subjects judged the sentences in the way we expected them to do, we report the mean accuracies (in %) and mean response latencies (in ms) for the correctly performed trials.

Subjects made 6.9% errors in the correct condition (COR), 9.9% in the condition with a mismatching dative (DAT), and 9.2% in the condition with a mismatching accusative (ACC). There was no main effect of VIOLATION [$F(2,30) = 1.02$, $P = 0.35$] and none of the single comparisons was significant [DAT vs. COR: $F(1,15) = 1.11$, $P = 0.31$; ACC vs. COR: $F(1,15) = 1.20$, $P = 0.29$; DAT vs. ACC: $F < 1$].

Mean response latencies were 507 ms in condition COR, 523 ms in DAT, and 535 ms in ACC. Again, there was no main effect of VIOLATION [$F(2,30) = 1.08$, $P = 0.35$] and none of the single comparisons was significant [DAT vs. COR: $F < 1$; ACC vs. COR: $F(1,15) = 1.52$, $P = 0.24$; DAT vs. ACC: $F < 1$]. Results show that subjects did not have problems in processing the sentences and that they were able to tell correct from incorrect sentences in the way we expected them to do.

4.2. Voltage-averaged ERP data

4.2.1. Descriptive results

Fig. 5 depicts the voltage-averaged ERPs at the critical word (verb, onset at 0 ms) in all three critical conditions for all 16 subjects at a subset of 9 electrodes. As can be clearly
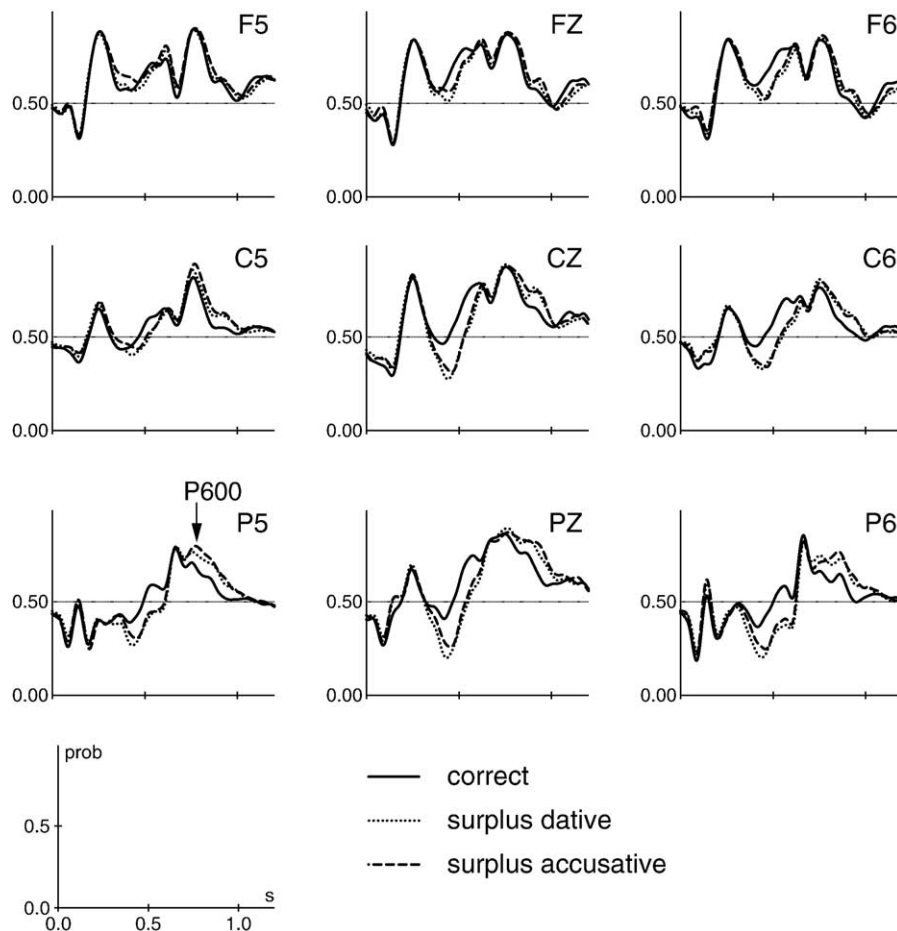
Fig. 7. Distribution of "2"s (relative frequency of trials with positive polarity) in each of the three critical conditions averaged over 16 subjects and over threshold range (II) (3.9 to 5.3 μV) from the onset of the critical item (verb, onset marked by vertical line) up to 1200 ms thereafter.

seen from the figure, both violation conditions show a negativity peaking at around 400 ms (N400) and a positivity peaking at around 750 ms (P600) compared to the correct condition. However, the patterns in the two violations are very much alike and seem to differ neither in the N400 nor the P600 time window.

4.2.2. Statistical results

4.2.2.1. Negativity time window (300 to 600 ms).
Over midline sites, we found a main effect of VIOLATION [$F(2,30) = 14.23$, $P < 0.01$], due to more negative going waveforms in both DAT [$F(1,15) = 28.32$, $P < 0.01$] and ACC [$F(1,15) = 14.90$, $P < 0.01$] compared to the correct condition. However, there was no difference between the two violation conditions ($F < 1$). An interaction VIOLATION × ELECTRODE [$F(4,60) = 11.10$, $P < 0.01$] was due to significant main effects of VIOLATION at electrodes CZ [$F(2,30) = 18.38$, $P < 0.01$] and PZ [$F(2,30) = 22.60$, $P < 0.01$] but only a marginal one at FZ [$F(2,30) = 2.92$, $P = 0.07$]. Sentences in the DAT condition were more negative going compared to correct sentences at FZ [$F(1,15) = 6.88$, $P < 0.05$], CZ [$F(1,15) = 33.64$, $P < 0.01$], and PZ [$F(1,15) = 41.84$, $P < 0.01$]. ACC

sentences differed from correct ones at CZ [$F(1,15) = 21.24$, $P < 0.01$] and PZ [$F(1,15) = 22.22$, $P < 0.01$], but not at FZ [$F(1,15) = 2.81$, $P = 0.11$]. However, the comparison between the two violation conditions did not reveal any significant differences at FZ ($F < 1$), CZ ($F < 1$), or PZ [$F(1,15) = 2.08$, $P = 0.26$].

Over lateral sites, we found a main effect of VIOLATION [$F(2,30) = 13.72$, $P < 0.01$], which was due to a negativity for both DAT [$F(1,15) = 25.43$, $P < 0.01$] and ACC [$F(1,15) = 11.80$, $P < 0.01$] compared to COR. However, there was no difference between the two violation conditions [$F(1,15) = 2.28$, $P = 0.23$]. Furthermore, we found an interaction VIOLATION × ROI [$F(10,150) = 9.78$, $P < 0.01$] which was due to main effects of VIOLATION in the right-anterior [$F(2,30) = 7.38$, $P < 0.01$], the left-central [$F(2,30) = 6.41$, $P < 0.01$], the right-central [$F(2,30) = 27.44$, $P < 0.01$], the left-posterior [$F(2,30) = 14.59$, $P < 0.01$], and the right-posterior [$F(2,30) = 33.49$, $P < 0.01$]. Mismatching datives induced negativities compared to correct sentences in each of these ROIs, as did the mismatching accusatives [right-anterior: DAT: $F(1,15) = 15.43$, $P < 0.01$, ACC: $F(1,15) = 5.85$, $P < 0.05$; left-central: DAT: $F(1,15) = 10.96$, $P < 0.01$, ACC: $F(1,15) = $

4.83, $P = 0.07$; right-central: DAT: $F(1,15) = 45.89$, $P < 0.01$, ACC: $F(1,15) = 32.56$, $P < 0.01$; left-posterior: DAT: $F(1,15) = 22.14$, $P < 0.01$, ACC: $F(1,15) = 15.29$, $P < 0.01$; right-posterior: DAT: $F(1,15) = 58.04$, $P < 0.01$, ACC: $F(1,15) = 46.64$, $P < 0.01$]. However, both violation conditions did not differ from each other in any of these ROIs (all $P > 0.21$).

4.2.2.2. Positivity time window (700 to 1000 ms). Over midline sites, we found a main effect of VIOLATION [$F(2,30) = 6.36$, $P < 0.01$] due to positivities in both the DAT [$F(1,15) = 7.42$, $P < 0.05$] and the ACC condition [$F(1,15) = 9.77$, $P < 0.05$] compared to the correct condition. Similar to the N400 time window, the two violation conditions did not differ from one another ($F < 1$).

Resolving an interaction VIOLATION × ELECTRODE [$F(4,60) = 5.04$, $P < 0.05$] revealed main effects of VIOLATION at CZ [$F(2,30) = 5.19$, $P < 0.05$] and PZ only [$F(2,30) = 10.46$, $P < 0.01$], but not at FZ [$F(2,30) = 1.60$, $P = 0.22$]. Both violation conditions differed from the correct condition at CZ [DAT vs. COR: $F(1,15) = 7.36$, $P < 0.05$; ACC vs. COR: $F(1,15) = 7.15$, $P < 0.05$] and at PZ [DAT vs. COR: $F(1,15) = 14.44$, $P < 0.01$; ACC vs. COR: $F(1,15) = 12.51$, $P < 0.01$]. Again, the violation conditions did not differ from one another at CZ or PZ (both $F < 1$).

Over lateral sites, there was a main effect of VIOLATION [$F(2,30) = 6.84$, $P < 0.01$] due to a positivity for both DAT [$F(1,15) = 7.12$, $P < 0.05$] and ACC [$F(1,15) = 11.08$, $P < 0.01$] compared to correct sentences. However, there was no difference between the two violation conditions ($F < 1$). There was no interaction VIOLATION × ROI [$F(10,150) = 1.98$, $P = 0.12$].

4.2.3. Summary

In sum, the statistical results of the ERPs confirm the visual inspection that both violation conditions induce an N400 followed by a P600 in comparison to the correct condition, whereas the two incorrect conditions differ from one another neither in the N400 nor the P600 time window.

4.3. Symbolic resonance analysis

4.3.1. Descriptive analysis

The optimal thresholds for the condition DAT versus ACC at the nine electrodes F5, FZ, F6, C3, CZ, C6, P5, PZ, and P6 plotted in Fig. 6 are given in Table 7 for the time windows 300 to 600 ms and 700 to 1000 ms.

According to the results given in Table 7, the *Reversi transformed* symbol distributions were averaged (I) over thresholds from 3.4 to 4.8 μV and (II) from 3.0 to 5.3 μV. Fig. 6 displays the relative frequencies of the symbol "0" (denoting negativity) of the *Reversi transformed* three-symbol encoded ERPs in all three conditions in the threshold range (I) averaged over all 16 subjects at a subset of 9 electrodes.

Table 7
Optimal thresholds for the comparison DAT versus ACC

Optimal threshold (μV) at channel	Time window (I) N400: 300 to 600 ms	Time window (II) P600: 700 to 1000 ms
F5	3.9	3.6
FZ	4.0	3.9
F6	4.1	3.0
C3	3.8	4.2
CZ	4.5	4.8
C6	3.8	3.6
P5	3.4	4.1
PZ	4.8	5.3
P6	3.9	4.5

As can be seen from the figure, both violation conditions DAT and ACC induce an N400 in the threshold range (I) compared to the correct condition. Moreover, the two violation conditions also differ from one another in this threshold range, in that the DAT condition shows a larger N400 compared to the ACC condition.

Fig. 7 displays the relative frequencies of the symbol "2" (denoting positivity) of the *Reversi transformed* three-symbol distributions in the threshold range (II) averaged in each of the three conditions over all 16 subjects at a subset of 9 electrodes.

As can be seen from the figure, both violation conditions DAT and ACC induce a P600 in the threshold range (II) compared to the correct condition. Additionally, the two violation conditions differ in this threshold range, too, showing a larger P600 for the ACC condition compared to the DAT condition.

4.3.2. Statistical analysis

4.3.2.1. Negativity time window (300 to 600 ms). For the threshold range (I) (3.4 to 4.8 μV) over midline sites, there was a main effect of VIOLATION [$F(2,30) = 110.23$, $P < 0.01$]. Compared to the correct condition, there was a higher proportion of the symbol "0" (meaning negativity) in the dative [$F(1,15) = 231.94$, $P < 0.01$] and in the accusative violation condition [$F(1,15) = 120.16$, $P < 0.01$]. Interestingly, there were also more "0"s for sentences with a mismatching dative compared to sentences with a mismatching accusative argument [$F(1,15) = 11.65$, $P < 0.01$]. Additionally, we found an interaction VIOLATION × ELECTRODE [$F(4,60) = 83.77$, $P < 0.01$], whose resolution revealed a main effect of VIOLATION at each of the three midline electrodes FZ [$F(2,30) = 17.22$, $P < 0.01$], CZ [$F(2,30) = 138.51$, $P < 0.01$], and PZ [$F(2,30) = 156.46$, $P < 0.01$]. Both violation conditions were more negative going compared to the correct condition at all midline electrodes: DAT vs. COR [FZ, $F(1,15) = 52.07$, $P < 0.01$; CZ, $F(1,15) = 264.46$, $P < 0.01$; PZ, $F(1,15) = 307.33$, $P < 0.01$] and ACC vs. COR [FZ, $F(1,15) = 6.25$ $P < 0.05$; CZ, $F(1,15) = 170.09$, $P < 0.01$; PZ, $F(1,15) = 171.01$, $P < 0.01$]. More importantly, we

observed that the symbolically encoded ERPs for the DAT condition were more negative than those for the ACC condition at all three midline sites: FZ [$F(1,15) = 8.35$, $P < 0.05$], CZ [$F(1,15) = 6.61$, $P < 0.05$], and PZ [$F(1,15) = 14.05$, $P < 0.01$].

Over lateral sites, there was a main effect of VIOLATION [$F(2,30) = 74.94$, $P < 0.01$] due to more coherence of the symbol "0" (denoting threshold crossing events with negative polarity) in both mismatching datives [$F(1,15) = 136.08$, $P < 0.01$] and accusatives [$F(1,15) = 65.67$, $P < 0.01$] in comparison to correct sentences. In the SRA, we additionally got a difference between the two violation conditions [$F(1,15) = 14.07$, $P < 0.01$] due to more "0"s in DAT compared to ACC. We found an interaction VIOLATION × ROI [$F(10,150) = 50.10$, $P < 0.01$] due to main effects of VIOLATION in all six ROIs [left-anterior: $F(2,30) = 6.41$, $P < 0.01$; right-anterior: $F(2,30) = 40.98$, $P < 0.01$; left-central: $F(2,30) = 26.76$, $P < 0.01$; right-central: $F(2,30) = 133.32$, $P < 0.01$; left-posterior: $F(2,30) = 76.27$, $P < 0.01$; right-posterior: $F(2,30) = 158.73$, $P < 0.01$]. Mismatching datives and accusatives induced negative coherent resonances compared to correct sentences in each of these ROIs [left-anterior: DAT: $F(1,15) = 8.14$, $P < 0.05$, ACC: $F(1,15) < 1$; right-anterior: DAT: $F(1,15) = 78.04$, $P < 0.01$, ACC: $F(1,15) = 38.43$, $P < 0.01$; left-central: DAT: $F(1,15) = 49.63$, $P < 0.01$, ACC: $F(1,15) = 17.93$, $P < 0.01$; right-central: DAT: $F(1,15) = 207.78$, $P < 0.01$, ACC: $F(1,15) = 204.53$, $P < 0.01$; left-posterior: DAT: $F(1,15) = 143.28$, $P < 0.01$, ACC: $F(1,15) = 64.40$, $P < 0.01$; right-posterior: DAT: $F(1,15) = 287.80$, $P < 0.01$, ACC: $F(1,15) = 156.98$, $P < 0.01$]. More interestingly, both violation conditions differed from each other in the all lateral ROIs [left-anterior: $F(1,15) = 12.78$, $P < 0.01$; right-anterior: $F(1,15) = 4.64$, $P = 0.07$; left-central: $F(1,15) = 9.24$, $P < 0.01$; right-central: $F(1,15) = 5.90$, $P < 0.05$; left-posterior: $F(1,15) = 12.09$, $P < 0.01$; right-posterior: $F(1,15) = 23.20$, $P < 0.01$]. These differences were due to more "0"s in DAT compared to ACC.

4.3.2.2. Positivity time window (700 to 1000 ms). The statistical analysis of the final distribution of "2"s in the late time window and in threshold range (II) (3.0 to 5.3 μV) at midline electrodes revealed a main effect of VIOLATION [$F(2,30) = 16.95$, $P < 0.01$] due to a greater number of positive ERP trials both in the dative violation condition [$F(1,15) = 16.85$, $P < 0.01$] and in the accusative violation condition [$F(1,15) = 23.09$, $P < 0.01$] compared to the correct condition. The two violation conditions did not differ from one another [$F(1,15) = 3.05$, $P = 0.15$]. Resolving an interaction VIOLATION × ELECTRODE [$F(4,60) = 8.52$, $P < 0.01$] revealed main effects of VIOLATION at FZ [$F(2,30) = 6.64$, $P < 0.01$], CZ [$F(2,30) = 14.29$, $P < 0.01$], and PZ [$F(2,30) = 24.33$, $P < 0.01$]. Both violation conditions differed from the correct condition at CZ [DAT: $F(1,15) = 19.54$, $P < 0.01$; ACC: $F(1,15) = 17.76$, $P < 0.01$] and at PZ [DAT: $F(1,15) =

28.72$, $P < 0.01$; ACC: $F(1,15) = 28.73$, $P < 0.01$]. At FZ, there was only a significant difference of the mismatching accusative condition against the correct condition [DAT: $F(1,15) = 2.04$, $P = 0.26$; ACC: $F(1,15) = 11.57$, $P < 0.01$] due to a higher proportion of "2"s in ACC. Both violation conditions DAT vs. ACC differed only at FZ [$F(1,15) = 6.79$, $P < 0.05$], but not at CZ [$F(1,15) = 1.14$, $P = 0.30$] or PZ ($F < 1$). There were more "2"s in ACC compared to DAT.

In the lateral ROIs, we observed a main effect of VIOLATION [$F(2,30) = 28.79$, $P < 0.01$]. Compared to the correct condition, there were more "2"s in both the mismatching dative [$F(1,15) = 29.23$, $P < 0.01$] and the mismatching accusative condition [$F(1,15) = 38.33$, $P < 0.01$]. By contrast, there was no global difference between the two violation conditions [$F(1,15) = 3.58$, $P = 0.12$]. An interaction VIOLATION × ROI [$F(10,150) = 8.33$, $P < 0.01$] was due to main effects of VIOLATION in all six ROIs [left-anterior: $F(2,30) = 10.57$, $P < 0.01$; right-anterior: $F(2,30) = 6.82$, $P < 0.01$; left-central: $F(2,30) = 36.37$, $P < 0.01$; right-central: $F(2,30) = 20.66$, $P < 0.01$; left-posterior: $F(2,30) = 27.02$, $P < 0.01$; right-posterior: $F(2,30) = 35.62$, $P < 0.01$]. In both violation conditions, we observed higher coherence of positive threshold crossing events in comparison to the correct condition in all ROIs except in right-anterior, where only the difference ACC vs. COR became significant [left-anterior: DAT: $F(1,15) = 6.76$, $P < 0.05$, ACC: $F(1,15) = 17.24$, $P < 0.01$; right-anterior: DAT: $F(1,15) = 3.99$, $P = 0.09$, ACC: $F(1,15) = 11.72$, $P < 0.01$; left-central: DAT: $F(1,15) = 42.98$, $P < 0.01$, ACC: $F(1,15) = 50.58$, $P < 0.01$; right-central: DAT: $F(1,15) = 25.54$, $P < 0.01$, ACC: $F(1,15) = 22.91$, $P < 0.01$; left-posterior: DAT: $F(1,15) = 29.45$, $P < 0.01$, ACC: $F(1,15) = 31.61$, $P < 0.01$; right-posterior: DAT: $F(1,15) = 45.15$, $P < 0.01$, ACC: $F(1,15) = 44.81$, $P < 0.01$]. The violation conditions differed from each other only marginally at frontal and left sites [left-anterior: $F(1,15) = 4.78$, $P = 0.07$; right-anterior: $F(1,15) = 3.44$, $P = 0.12$; left-central: $F(1,15) = 5.37$, $P = 0.05$; right-central: $F(1,15) < 1$; left-posterior: $F(1,15) = 3.51$, $P = 0.12$; right-posterior: $F(1,15) < 1$], these differences being due to more "2"s in ACC.

4.3.3. Summary

Summing up, the statistical results of the symbolic resonance analysis of ERPs confirm the results from the voltage averaging analysis in that both violation conditions induce an N400 followed by a P600 in comparison to the correct condition. In addition, the symbolic resonance analysis was able to reveal a difference between the two incorrect conditions, too. We observed a larger N400 component for the mismatching dative condition compared to the mismatching accusative and a marginally larger P600 component at frontal electrode sites for the mismatching accusative condition in comparison with the mismatching dative condition.

5. Discussion

Previous ERP studies have shown that sentences in which the number of NP arguments does not match the number of arguments specified in the verb's lexical entry lead to a biphasic N400–P600 response [11,15]. This pattern, which had been found with both intransitive [15] and transitive constructions [11], has been replicated with ditransitive constructions in the present study. As would have been expected, sentences with verbs that could not integrate either an accusative object or a dative object elicited a biphasic N400–P600 pattern in the ERP. In the context of argument structure violations, the N400 has been interpreted to signal the thematic–semantic integration problem, whereas the P600 has been argued to reflect the fact that the syntactic structure is not licensed by the lexicon information of the specific verb [10,15]. This view supports the theoretical characterization of argument structure information as an interface between syntactic (i.e., syntactic structure) as well as semantic (i.e., propositional) aspects of sentences in a crucial way.

In the voltage average analysis, there was no difference between the two incorrect conditions, that is, between sentences in which a dative (i.e., indirect) object cannot be integrated compared to sentences with verbs which could not take an accusative (i.e., direct) object argument. However, such a difference was found in the symbolic resonance analysis (SRA), a specific type of data analysis that – to our knowledge – was applied to a neurophysiological data set here for the first time. The SRA not only replicated the voltage average findings of differences between each of the violation conditions to the correct condition, but also revealed that a surplus dative construction leads to a stronger "N400" (i.e., higher proportion of symbol "0") and a weaker "P600" (i.e., lower proportion of symbol "2") in the symbolic resonance analysis compared to a condition with a surplus accusative. Since the N400 has been seen as a marker of semantic–thematic integration problems, whereas the P600 has been interpreted as reflecting syntactic ill-formedness (cf. [10]), one might conclude that an incorrect dative argument induces more semantic and less syntactic processing effort compared to an incorrect accusative argument. Although it should be kept in mind that the "N400 = semantics" versus "P600 = syntax" distinction is only a rule of thumb, it would not be implausible with respect to the present data to assume that a surplus dative violation induces a weaker syntactic mismatch but a stronger semantic mismatch correlate compared to the mismatching accusative condition. This would make sense insofar as these two cases in German exhibit different characteristics: Transitive nominative–accusative verbs in German allow for an additional dative expressing who is profiting from what is entailed in the proposition (i.e., expressing a *beneficent*). Adding such a benefactive dative seems to be in principal (syntactically) possible in these types of verbs, but is semantically

restricted [34,37]. One could therefore speculate that if such a construction is rejected by a speaker, this could be for reasons which are primarily semantic in nature. It might suggest that the language processing system spends more effort in the semantic analysis, in order to judge whether the free dative is semantically acceptable. By contrast, a mismatching accusative argument leads to a violation which might be more principal in nature, that is, which might also include syntactic restrictions, seeing that adding an accusative is not possible at all. One might dispute the possibility of a free accusative in German also exists in sentences such as (15).

(15). Anna reparierte das Motorrad **den ganzen Tag**.
*Anna repaired the motorcycle **the whole day**.*

We do not think that this possibility is a problem for our argument for the following reasons: Free accusatives such as "den ganzen Tag" (*the whole day*) are extremely limited with respect to lexicon, semantics, and syntax (that is, they are limited to nouns expressing a duration). For example, leaving out the adjective already results in an argument structure violation, see (16).

(16). *Anna reparierte das Motorrad **den Tag**.
*Anna repaired the motorcycle **the day.***

Furthermore, as (15) also shows, the accusative can be added in a sentence that already has a direct object in accusative case, which clearly demonstrates that it is not an argument of the verb at all. The same is shown by the fact that the free accusative is not involved in processes of argument reassignment as in passive constructions, see (17).

(17). *Der ganze Tag wurde von Anna repariert.
the whole day was repaired by Anna.

Benefactive datives in German, by contrast, meet all requirements for an argument status (cf. [9,12] and Footnote 1).

We admit that the above interpretation of the difference between extra-datives and extra-accusatives might be tentative. In any case, more theoretical and especially empirical work has to be done in other languages with morphological case marking as well, in order to further elucidate how this central issue of sentence processing, namely, argument interpretation, is achieved by our brains. What we can clearly say, however, is that our finding of differences between extra-datives and extra-accusatives in processing is limited to a specific type of data analysis that we present here for the first time. As we have shown in the present paper, the symbolic resonance analysis, in which the intrinsic noise of the EEG is utilized for enhancing the ERP signal, offers a new way to distinguish ERP reflections where the conventional voltage average technique fails. Thus, the symbolic resonance analysis is more sensitive than conventional methods, complementing these techniques. This offers a promising starting point for revealing fine-

grained differences in a range of data which the traditional analysis techniques are too insensitive to unveil.

Acknowledgments

References

[1] R.H. Baayen, R. Piepenbrock, H. van Rijn, The CELEX Lexical Database (CD-ROM), Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 1993.

[2] R. Badii, A. Politi, Complexity. Hierarchical Structures and Scaling in Physics, Cambridge University Press, Cambridge, UK, 1997.

[3] P. beim Graben, Estimating and improving the signal-to-noise ratio of time series by symbolic dynamics, Phys. Rev. E 64 (2001) 051104.

[4] P. beim Graben, J. Kurths, Detecting subthreshold events in noisy data by symbolic dynamics, Phys. Rev. Lett. 90 (10) (2003) 100602.

[5] P. beim Graben, S. Frisch, Is it positive or negative? On determining ERP components, IEEE Trans. Biomed. Eng. 51 (2004) 1374–1382.

[6] P. beim Graben, J.D. Saddy, M. Schlesewsky, J. Kurths, Symbolic dynamics of event-related brain potentials, Phys. Rev. E 62 (4) (2000) 5518–5541.

[7] I. Bornkessel, M. Schlesewsky, A.D. Friederici, Beyond syntax: language-related positivities reflect the revision of hierarchies, NeuroReport 3 (2002) 361–364.

[8] E. Callaway, R.A. Halliday, Evoked potential variability: effects of age, amplitude and methods of measurement, Electroencephalogr. Clin. Neurophysiol. 34 (1973) 125–133.

[9] P. Eisenberg, Grundriß Der Deutschen Grammatik, 3. Aufl., Metzler, Stuttgart/Weimar, 1994.

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

[11] A.D. Friederici, S. Frisch, Verb argument structure processing: the role of verb-specific and argument-specific information, J. Mem. Lang. 43 (2000) 476–507.

[12] S. Frisch, Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen, MPI of Cogn. Neurosci., Leipzig, 2000.

[13] S. Frisch, M. Schlesewsky, The N400 reflects problems of thematic hierarchizing, NeuroReport 12 (2001) 3391–3394.

[14] S. Frisch, M. Schlesewsky, The multidimensionality of case as revealed by ERPs, J. Cogn. Neurosci., Suppl. (2003) 35.

[15] S. Frisch, A. Hahne, A.D. Friederici, Word category and verb-argument structure information in the dynamics of parsing, Cognition 91 (2004) 191–219.

[16] S. Frisch, P. beim Graben, M. Schlesewsky, Parallelizing grammatical functions: P600 and P345 reflect different cost of reanalysis, Int. J. Bifurc. Chaos 14 (2004) 531–549.

[17] Y. Grodzinsky, The neurology of syntax: language use without Broca's area, Behav. Brain Sci. 23 (2000) 1–71.

[18] B.-L. Hao, Symbolic dynamics and characterization of complexity, Physica D 51 (1991) 161–176.

[19] G. Helbig, J. Buscha, Deutsche Grammatik, Langenscheidt/Verl. Enzykl., Leipzig/Berlin/München, 1991.

[20] H. Huynh, L.S. Feldt, Conditions under which the mean square ratios in repeated measurement designs have exact F-distributions, J. Am. Stat. Assoc. 65 (1970) 1582–1589.

[21] G. Keppel, Design and analysis, Prentice Hall, Upper Saddle River, 1991.

[22] D. Lehmann, Multichannel topography of human alpha EEG fields, Electroencephalogr. Clin. Neurophysiol. 31 (1971) 439–449.

[23] D. Lind, B. Marcus, Symbolic Dynamics and Coding, Cambridge University Press, Cambridge, UK, 1995.

[24] S. Makeig, M. Westerfield, T.-P. Jung, S. Enghoff, J. Townsend, E. Courchesne, T.J. Sejnowski, Dynamic brain sources of visual evoked responses, Science 295 (2002) 690–694.

[25] F. Moss, D. Pierson, D. O'Gorman, Stochastic resonance: tutorial and update, Int. J. Bifurc. Chaos 4 (1994) 1383–1397.

[26] F. Moss, L.M. Ward, W.G. Sannita, Stochastic resonance and sensory information processing: a tutorial and review of application, Clin. Neurophysiol. 115 (2004) 267–281.

[27] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh Inventory, Neuropsychologia 9 (1971) 97–113.

[28] L. Osterhout, On the brain responses to syntactic anomalies: manipulations of word position and word class reveal individual differences, Brain Lang. 59 (1997) 494–522.

[29] L. Osterhout, P.J. Holcomb, D.A. Swinney, Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing, J. Exp. Psychol. Learn. Mem. Cogn. 20 (1994) 786–803.

[30] B. Primus, Cases and Thematic Roles, Niemeyer, Tübingen, 1999.

[31] C.E. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1949.

[32] R.D. Van Valin, R.J. LaPolla, Syntax, Cambridge University Press, Cambridge, 1997.

[33] S. Wechsler, The Semantic Basis of Argument Structure, CSLI publications, Stanford, 1995.

[34] H. Wegener, Der Dativ- ein struktureller Kasus? in: G. Fanselow, S.W. Felix (Eds.), Strukturen und Merkmale Syntaktischer Kategorien, Narr, Tübingen, 1991, pp. 70–103.

[35] F.Y. Wu, The Potts model, Rev. Mod. Phys. 54 (1982) 235–268.

[36] D. Wunderlich, Über die Argumente des Verbs, Ling, Berichte 97 (1985) 183–228.

[37] D. Wunderlich, Diathesen, in: J. Jacobs, A.v. Stechow, W. Sternefeld, T. Vennemann (Eds.), Syntax: Ein Internationales Handbuch Zeitgenössischer Forschung, vol. 1, de Gruyter, Berlin, 1993, pp. 730–747.