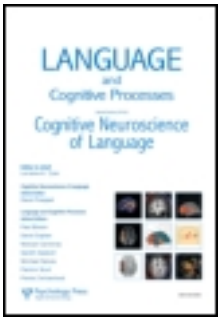

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Processing flexible form-to-meaning mappings: Evidence for enriched composition as opposed to indeterminacy

Dietmar Roehm [a], Antonella Sorace [b c] & Ina Bornkessel-Schlesewsky [d]

[a] Department of Linguistics, University of Salzburg, Salzburg, Austria

[b] Department of Linguistics, University of Edinburgh, Edinburgh, UK

[c] CASTL, University of Tromsø, Tromsø, Norway

[d] Department of Germanic Linguistics, University of Marburg, Marburg, Germany

Version of record first published: 09 Oct 2012.

To cite this article: Dietmar Roehm, Antonella Sorace & Ina Bornkessel-Schlesewsky (2012): Processing flexible form-to-meaning mappings: Evidence for enriched composition as opposed to indeterminacy, Language and Cognitive Processes, DOI:10.1080/01690965.2012.712143

To link to this article: http://dx.doi.org/10.1080/01690965.2012.712143

LANGUAGE AND COGNITIVE PROCESSES, 2012, iFirst, 1–31

Psychology Press
Taylor & Francis Group

Processing flexible form-to-meaning mappings: Evidence for enriched composition as opposed to indeterminacy

Dietmar Roehm[1], Antonella Sorace[2,3], and Ina Bornkessel-Schlesewsky[4]

[1]Department of Linguistics, University of Salzburg, Salzburg, Austria
[2]Department of Linguistics, University of Edinburgh, Edinburgh, UK
[3]CASTL, University of Tromsø, Tromsø, Norway
[4]Department of Germanic Linguistics, University of Marburg, Marburg, Germany

Sometimes, the relationship between form and meaning in language is not one-to-one. Here, we used event-related brain potentials (ERPs) to illuminate the neural correlates of such flexible syntax–semantics mappings during sentence comprehension by examining split-intransitivity. While some ("rigid") verbs consistently select one auxiliary, other ("flexible") verbs do not. German speakers read sentences including rigid and flexible verbs and the auxiliaries *haben* (HAVE) and *sein* (BE). In an additional condition, flexible verbs were presented with a telicity-inducing prefix that led them to select BE via the verb-prefix combination. Auxiliary selection violations engendered an N400-late positivity response for both rigid and prefixed verbs, thus suggesting that the processing system sets up an auxiliary-based expectation for particular verb classes. For unprefixed flexible verbs, average ERPs did not show differential effects of auxiliary choice. However, additional mixed-effects analyses for these verbs including by-participant and by-item acceptabilities as covariates revealed modulations of the N400 and late positivity for HAVE-trials via by-item and by-participant acceptabilities. We argue that the N400 reflects the degree of match between auxiliary choice and the lexical class of the verb, while two distinct positivity effects correlate with (1) a well-formedness categorisation and (2) the degree to which individual participants engage in enriched composition in order to render a flexible (telic) verb compatible with a particular auxiliary choice (HAVE). These results indicate that the gradient behaviour of flexible verbs is not due to indeterminacy, but rather to a higher propensity for enriched composition that results in the coercion of aspectual specification.

Keywords: Language comprehension; Syntax-semantics interface; Split intransitivity; Enriched composition; Underspecification; Event-related brain potentials.

Correspondence should be addressed to Ina Bornkessel-Schlesewsky, Department of Germanic Linguistics, University of Marburg, Wilhelm-Roepke-Strasse 6A, 35032 Marburg, Germany. E-mail: iboke@staff.uni-marburg.de

Parts of the research reported here were performed while D.R. and I.B.S. were at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. In addition, parts of this research were supported by the Hessian Ministry for Science and Arts (LOEWE-programme) as part of the research group "Stability and change in language: theoretical and empirical approaches". We would like to thank Matthias Schlesewsky for helpful discussions and Bettina Gutsche for assistance in the preparation of the experimental materials as well as in data preprocessing.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2012.712143

Language is an extremely efficient medium for human communication because it allows us to express and understand complex meanings in real time. This astounding efficacy is guaranteed, in part, by the system's ability to exploit systematic correspondences between form and meaning. For example, in many languages, the type of auxiliary (*be* vs. *have*) provides constraining information about whether an upcoming intransitive verb is unaccusative (1a) or unergative (1b) and, accordingly, whether the subject of the sentence is likely to be a Patient or nonvolitional Causer (1a) or an Agent (1b) (Dowty, 1991; Van Valin, 1990). This phenomenon is known as "split intransitivity".[1]

(1) a. Il bambino è arrivato/*ha arrivato in ritardo ITALIAN
The child is/has arrived late

b. I bambini hanno giocato/*sono giocati tutto il pomeriggio
The children have played/are played the whole afternoon

In certain cases, however, there are multiple possibilities as to how the form-to-meaning mapping could be undertaken, that is, flexibility with regard to the possible meanings that could be associated with a given form and vice versa. In the domain of split intransitivity, for example, there are verbs that can select either BE or HAVE depending on the characteristics of the predicate (see Levin & Rappaport Hovav, 1995, for a full discussion). Under these circumstances, an auxiliary thus does not provide the language processing system with an unambiguous indication of verb meaning/semantic role of the subject.

The aim of the present study was to investigate how the human language comprehension system deals with flexibility versus rigidity in the form-to-meaning mapping and to examine the neural correlates of processing flexible mappings. In the next section, we briefly describe the factors governing flexibility versus rigidity in the syntax–semantics mapping for intransitive verbs, before reporting an event-related brain potential (ERP) study that manipulated these factors.

Split intransitivity: Rigidity versus flexibility at the syntax-semantics interface

The selection of perfective auxiliaries *avere/haben* ("have") and *essere/sein* ("be") with intransitive verbs in languages such as Italian and German is an ideal testing ground to study variation at the syntax–semantics interface. Auxiliary selection has traditionally been regarded as a syntactic diagnostic of unaccusativity (Burzio, 1986; Perlmutter, 1978). The Unaccusative Hypothesis states that different types of intransitive verbs are associated with distinct syntactic configurations. Recent research, however, has shown that the notion of a binary distinction between unaccusative and unergative verbs is difficult to maintain, since not all verbs of a certain type behave in the same way: preferences for one auxiliary or the other systematically vary in strength depending on the aspectual type of the main verb and the context in which the auxiliary-verb combination appears (Cennamo & Sorace, 2007; Keller & Sorace, 2003; Legendre, 2007; Legendre & Sorace, 2003; Sorace, 2000, 2004, to appear, among others).

[1] Note that, in addition to auxiliary selection, split intransitivity also correlates with other syntactic properties (Burzio, 1986; Perlmutter, 1978). For the purpose of the present paper, however, we will focus on auxiliary selection.

Systematic verb variability poses a problem for the Unaccusative Hypothesis and theoretical linguistic research has sought to address this issue by focusing on the complex mappings between a lexical-semantic level of representation and the level of syntactic structure (Levin & Rappaport Hovav, 2005; Ramchand, 2008). Two broad types of solutions have been proposed. "Projectionist" approaches enrich the lexical entry of verbs with fine-grained semantic specifications, which project to the syntax via a complex system of linking rules. "Constructional" approaches, on the other hand, assume "bare" lexical entries that are free to project onto enriched syntactic configurations, which in turn determine interpretation. [For a detailed discussion of this distinction, see Levin and Rappaport Hovav (2005).] Both these accounts have limitations: the projectionist view allows for too little variation, because of the deterministic nature of its linking rules, whereas the constructionist view allows too much variation, because it lacks a mechanism that rules out impossible mappings. In other words, neither type of account—in its pure form—explains the fact that the system is both rigid and flexible (McFadden, 2007).[2] [Note, however, that some recent versions of the constructionist approach (e.g. Borer, 2005; Ramchand, 2008) do not completely empty lexical entries of featural specification, so that some verbs are preferentially associated with particular syntactic configurations.] The issue of rigidity versus flexibility in the syntax-semantics mapping for intransitive verbs has been highlighted in particular by work by Sorace and colleagues (see e.g. Keller & Sorace, 2003; Sorace, 2000, 2004, to appear), who have proposed that intransitive verbs are organised in a split intransitivity hierarchy (SIH; see Figure 1).

The array of verb classes represented on the SIH reduces to two key factors, the interaction of which affects the syntax of split intransitivity and creates gradient auxiliary preferences: "telic change" at the core of unaccusativity and "atelic nonmotional activity" at the core of unergativity. The closer to the core a verb is, the more determinate its syntactic status as either unaccusative or unergative. Sensitivity to contextual or compositional factors also correlates with the distance of a verb from the core: verbs that are stative and nonagentive are the most indeterminate and therefore the most susceptible to syntactic alternations. The hierarchy has received support from an increasing number of studies in typologically

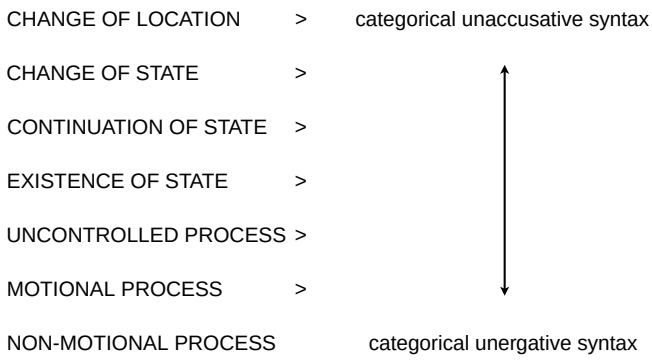

Figure 1. The split-intransitivity hierarchy (SIH).

[2]Note that the arguments against pure projectionist accounts also speak against polysemy (i.e. the assumption of more than one meaning for the same verb) as a general means of modelling gradient auxiliary selection. A polysemy-based account predicts that auxiliary selection alternations for individual verbs should be exceptional; however, they are in fact quite common (see McFadden, 2007, for details).

diverse languages, including Basque, French, Catalan, Chinese, Croatian, Dutch, German, Italian, Japanese, Paduan, Sardinian, Spanish, Turkish and, in addition, some sign languages (Sorace, to appear). Further converging support for gradience in auxiliary selection stems from behavioural experiments using acceptability judgments (Bard, Robertson, & Sorace, 1996; Keller & Sorace, 2003), developmental findings (Montrul, 2005; Sorace, 1993, 1995), and more recently from eye-tracking findings on online processing (Bard, Frenck-Mestre, & Sorace, 2010).

Processing flexible syntax–semantics mappings: Research questions

The central aim of the present study was to examine how the neural language comprehension system processes flexible syntax–semantics mappings, that is, structures in which a verb can acceptably combine with either *have* or *be*. While previous studies have revealed that these types of constructions produce gradient behaviour (e.g. acceptability ratings that are intermediate between those of sentences with categorical verbs, Keller & Sorace, 2003), they were not suited to revealing the mechanisms that result in this behaviour. In this regard, there appear to be two broadly possible scenarios[3]:

(2) a. *Indeterminacy Hypothesis*. Flexible verbs are semantically and/or syntactically indeterminate (underspecified).
b. *Enrichment Hypothesis*. Flexible verbs have a higher propensity for enriched composition than categorical verbs.

According to the Indeterminacy Hypothesis, flexible verbs are simply not specified for one configuration as opposed to the other. This underspecification could either be syntactic (assuming that unaccusativity is a syntactic phenomenon, see e.g. Rosen; Levin & Rappaport Hovav, 1995) or semantic in nature (assuming that unaccusativity is semantically conditioned, e.g. Legendre, 2007; Sorace, 2000; Van Valin, 1990). From this perspective, a flexible verb should show no effects of priming or prediction as a function of a preceding auxiliary. The Enrichment Hypothesis, by contrast, posits that flexible verbs are not—or at least not completely—underspecified, but that, in contrast to rigid verbs, they more readily allow for enriched composition that shifts their meaning to one that fits the auxiliary with which they are combined. This type of enrichment can be envisaged, for example, as a coercion operation that shifts the semantic type of an element (for a review from a psycholinguistic perspective, see Pylkkänen & McElree, 2006). By contrast, it can also be viewed as a pragmatic operation (e.g. Dölling, 1995; for discussion from a psycholinguistic perspective, see Brennan & Pylkkänen, 2008). The Enrichment Hypothesis thus predicts auxiliary-induced asymmetries in the online processing of flexible verbs, but does not commit a priori to whether these processes are semantic or pragmatic in nature.

The present ERP study sought to contrast these two hypotheses by examining the electrophysiological processing signatures of flexible verbs following either *have* or *be* in a sentence context and to contrast these with the processing signatures of auxiliary selection violations in rigid verbs.

[3] Note that, in sketching out these possible alternative processing strategies, we do not intend for a direct association with existing grammatical accounts of the syntax–semantics interface (see the *Split intransitivity: Rigidity versus flexibility at the syntax–semantics interface* section), since none of these were intended as processing models.

Previous psycholinguistic and neurolinguistic findings on indeterminacy versus enrichment

Questions regarding enrichment or indeterminacy have been examined in a range of psycholinguistic and neurolinguistic studies on complement/aspectual coercion. Complement coercion involves a shift from an entity to an event as in *The boy began the book*, while aspectual coercion involves a shift in the lexical aspect of a verb (e.g. from a punctual to an iterative meaning as in *Peter jumped for an hour*). For complement coercion, there has been some discussion with regard to whether it engenders online processing cost (e.g. McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; Traxler, Pickering, & McElree, 2002), which is compatible with a lexical mismatch followed by a coercion operation, or not (de Almeida, 2004; but see Pickering, McElree, & Traxler, 2005). Across these various studies, evidence appears to be accruing in favour of a coercion cost, as also supported by more recent experiments using neurocognitive methods (EEG: Baggio, Choma, van Lambalgen, & Hagoort, 2010; Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010; MEG: Pylkkänen & McElree, 2007; fMRI: Husband, Kelly, & Zhu, 2011). However, this still leaves open the question of whether the coercion operation is semantic or pragmatic in nature—an issue that is very difficult to determine empirically. We shall return to this issue in the *Discussion* section.

The literature has seen a somewhat similar debate with regard to aspectual coercion, with some authors reporting (behavioural) processing costs for the required shift to an iterative reading of a punctual verb (e.g. Piñango, Winnick, Ullah, & Zurif, 2006), while others have failed to observe such effects (Pickering, McElree, Frisson, Chen, & Traxler, 2006). On the basis of their findings, Pickering et al. (2006) proposed that aspectual coercion and complement coercion may allow for underspecification to different degrees, perhaps due to the fact that only complement coercion requires a type-shift for successful interpretation. Recently, however, this assumption has been challenged; using strongly punctual verbs, Brennan and Pylkkänen (2008) demonstrated aspectual coercion costs in self-paced reading and also observed increased neural activation for their coercion conditions using MEG (see below for further details on their results). In the present study, we used EEG to investigate questions of indeterminacy versus enriched composition in a somewhat similar phenomenon in German: like aspectual coercion, auxiliary selection with gradient verbs is tied to the telicity of the verbs under consideration and to how it is represented and used by the processing system.

Our experimental design and more concrete predictions for the ERP findings will be discussed in the following section.

The present study: Design and hypotheses

In order to examine the processing of flexible syntax–semantics mappings in German, the present study employed four critical verb types (see Table 1). These verbs were selected on the basis of a previous behavioural study by Keller and Sorace (2003)[4], which demonstrated a consistent preference for BE for verbs of change of location

[4] In Keller and Sorace's (2003) Experiment 2, the four verb classes examined here yielded the following acceptability ratings in a magnitude estimation study when combined with HAVE and BE in a sentence context (normalised, log-transformed judgements):

(i) CH-LOC: − 0.4 (HAVE) vs. 0.35 (BE)
(ii) CH-STATE-UN: 0.05 (HAVE) vs. 0.1 (BE)
(iii) CH-STATE-PRE: − 0.35 (HAVE) vs. 0.3 (BE)
(iv) CON-PROC: 0.35 (HAVE) vs. –0.45 (BE)

TABLE 1
Example sentences for the critical conditions in the present study as well as mean acceptability ratings and mean accuracy rates for the probe detection task

Condition *Verb class*	*Example*	*Acceptability (%)*		*Probe detection accuracy (%)*	
		HABEN	*SEIN*	*HABEN*	*SEIN*
CH-LOC	Die Bergsteigerin ist/*hat vorsichtig aufgestiegen. The mountaineer is/has carefully climbed "The mountaineer climbed carefully."	1.9 (3.0)	93.5 (6.0)	97.5 (3.1)	98.9 (2.2)
CH-STATE-UN	Die Dose ist/hat sofort gerostet. The tin is/has immediately rusted "The tin rusted immediately."	61.0 (17.0)	65.0 (19.6)	98.8 (1.8)	98.4 (2.2)
CH-STATE-PRE	Das Auto ist/*hat langsam verrostet. The car is/has slowly corroded "The car corroded slowly."	6.0 (4.4)	88.6 (8.6)	97.5 (2.0)	98.7 (1.7)
CON-PROC	Die Lehrerin *ist/hat dauernd geredet. The teacher is/has constantly talked "The teacher talked constantly."	94.3 (5.7)	1.2 (2.3)	99.0 (2.0)	97.3 (2.4)

Notes: Standard deviations (by participants) are given in parentheses.
CH-LOC, change of location; CH-STATE-UN, unprefixed change of state; CH-STATE-PRE, prefixed change of state; CON-PROC, controlled nonmotional process.

(CH-LOC) and for HAVE with verbs of controlled nonmotional process (CON-PROC). Unprefixed change of state verbs (CH-STATE-UN) exhibited flexibility with regard to auxiliary selection, while the addition of a telicity-inducing prefix (CH-STATE-PRE) led the same verbs to show a strong preference for BE. Thus, the present study included core unaccusative verbs (CH-LOC), core unergative verbs (CON-PROC), and an intermediate class that is not inherently specified for telicity (CH-STATE-UN). Prefixed change of state verbs (CH-STATE-PRE) were additionally included in order to investigate possible differences between lexical (CH-LOC) and morphosyntactic specification of telicity (via prefixation; CH-STATE-PRE).

Our hypotheses for the ERP data at the position of the critical clause-final verb (underlined in Table 1) were as follows:

For auxiliary selection violations with rigid verbs (CH-LOC, CON-PROC), we expected to observe a biphasic N400-late positivity pattern in comparison to correct control conditions. An N400 modulation was expected for two reasons. On the one hand, N400 amplitude decreases when a word is lexically preactivated by the preceding sentence context (Lau, Phillips, & Poeppel, 2008). Assuming that HAVE and BE serve to preactivate verbs with compatible semantic properties (atelic nonmotional activity vs. telic change), incompatible auxiliary-verb combinations involve lower lexical preactivation of the verb, hence leading to an N400 increase. On the other hand, N400 amplitudes have been shown to increase with increasing effort in the syntax–semantics linking, for example, for mismatches between word order and verb class in grammatical sentences (e.g. Bornkessel, McElree, Schlesewsky, & Friederici, 2004; Bornkessel-Schlesewsky, & Schlesewsky, 2008b) or between case marking and grammatical aspect (Choudhary, Schlesewsky, Roehm, & Bornkessel-Schlesewsky, 2009). Thus, N400 modulations were predicted both from the perspective of lexical preactivation (or the absence of it) and with regard to linking requirements. We additionally expected to observe late positivity effects for violations versus controls, since these typically occur in unacceptable sentences (particularly in experiments employing an acceptability judgement task), possibly reflecting well-formedness categorisation (Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011; Bornkessel-Schlesewsky et al., 2011) or "conflict monitoring" (Kolk, Chwilla, van Herten, & Oor, 2003; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010, among others).

For auxiliary selection violations involving the compositional specification of telicity, we have no concrete predictions. However, we included this verb class in order to examine whether compositionally specified telicity would engender qualitatively or quantitatively different effects to lexically specified telicity. Whether or not this is the case will help to shed light on the functional interpretation of the effects that we observe for the core verb classes (CH-LOC, CON-PROC), particularly on the question of whether any observed N400 effects are conditioned by lexical preactivation or by linking. If lexical preactivation is key, we should observe a difference between the effects for lexically specified and compositionally determined auxiliary choices.

For the flexible verbs (CH-STATE-UN), we can formulate different predictions based on the Indeterminacy Hypothesis and Enrichment Hypothesis, respectively. In this context, the Indeterminacy Hypothesis is essentially the null hypothesis, since it does not lead us to expect any ERP differences between CH-STATE-UN verbs in the context of HAVE and BE. The alternative hypothesis, The Enrichment Hypothesis, does predict an asymmetry between auxiliaries, with increased coercion costs expected for HAVE versus BE. Even though the CH-STATE-UN verbs used here were not strongly specified for telicity at the lexical level, a change of state does imply telicity at some level; these verbs are therefore closer to the "BE end" of the SIH (see Figure 1

and, for further discussion, Sorace, 2000).[5] When used with HAVE, they must be coerced to a process (i.e. activity) reading in order for the sentence to be acceptable. Processes of enriched composition have been shown to correlate with N400 and late positivity effects in previous electrophysiological studies (Baggio et al., 2010; Kuperberg et al., 2010; Schumacher, 2011). Interestingly, the type of effect observed varies with the type of coercion required: whereas complement coercion engenders increased N400 effects (Baggio et al., 2010; Kuperberg et al., 2010), reference transfer (i.e. using a salient property of an entity to refer to that entity, as in *The ham sandwich is sitting at Table 20*, from Nunberg, 1979) elicits an increased late positivity effect (Schumacher, 2011). The qualitative difference between the two types of enriched composition could be due to the increased involvement of discourse-pragmatic properties in the second case, with reference transfer crucially requiring a supporting context (see Schumacher, 2011, for discussion). However, since complement coercion and reference transfer differ along a number of dimensions, the functional distinction between the "enrichment N400" and the "enrichment positivity" is difficult to concretise at present. Functionally, the coercion which we claim might be required in the processing of flexible verbs appears closest to "aspectual coercion", which has been shown to share neuromagnetic correlates with complement coercion, namely activation of the "anterior midline field" (AMF) in ventromedial prefrontal cortex (for aspectual coercion: Brennan & Pylkkänen, 2008; for complement coercion: Pylkkänen & McElree, 2007).[6] On the basis of this result, we might expect to observe N400 modulations for HAVE versus BE with flexible verbs in the present study. However, late positivity modulations also cannot be fully ruled out.

Finally, if the enrichment account is correct, we could expect to observe a correlation between ERP effects and acceptability ratings at the subject and item level. As already noted above, work within the context of the SIH has demonstrated variability across verbs with regard to the flexibility versus rigidity of auxiliary selection. From this perspective, it appears natural that, even within a particular class, there may be differences with regard to how accessible coercion is for a particular verb. Likewise, we might expect different individuals to be more or less likely to apply coercive processes with flexible verbs (see St. George, Mannes, & Hoffman, 1997 for correlations between ERP responses and interindividual differences in inferencing abilities, and Nieuwland, Ditman, & Kuperberg, 2010, for a recent demonstration of interindividual differences in pragmatic abilities and a correlation with ERP patterns). (Note that, since previous behavioural studies of gradient auxiliary choice did not differentiate between item- and participant-based sources of gradience, there is no clear hypothesis as to whether one of these sources should be particularly important as opposed to the other. However, both appear plausible a priori.) In order to examine these expectations, we performed additional analyses of our ERP data in order to determine whether the effects observed correlate with item- and/or subject-based

[5] In fact, they also tend to prefer BE in other languages, though not as strongly as verbs that are inherently specified for telicity. In Italian, for example, these verbs have a noncategorical preference for essere, 'be' (Sorace, 2000): La panchina è arrugginita/??ha arrugginito sotto la pioggia ("The bench is/has rusted under the rain.").

[6] To complicate matters, however, the AMF activation observed for complement coercion by Pylkkänen and McElree (2007) was measurably distinct from activation engendered by semantic violations (i.e. violations which typically modulate the N400 in ERP studies). This appears to contradict the observation that complement coercion elicits N400 effects (Baggio et al., 2010; Kuperberg et al., 2010). However, since EEG and MEG activity are not directly comparable, we base our predictions more strongly on previous ERP findings. We will return to possible links to the MEG studies in the *Discussion* section.

acceptability measures. Potential correlations of this type will serve to further illuminate the functional significance of the observed ERP effects: while in mismatches arising from violations of expectations, higher ERP amplitudes generally correlate with lower acceptabilities, the successful application of a coercion operation should be expected to increase acceptability on that particular trial. In the latter scenario, we might therefore expect to observe that higher acceptability correlates with higher ERP amplitudes. Crucially, these two scenarios are not mutually exclusive in the sense that they could be reflected in different parts of the ERP signal (see, for example, Burkhardt, 2006, for the observation that referents requiring bridging inferences lead to reduced N400 effects in comparison to new referents, while at the same time engendering increased late positivity responses in comparison to contextually given referents).

MATERIALS AND METHODS

Participants

Thirty-two monolingually raised native speakers of German (students of the University of Leipzig; 17 women; mean age 23.97, range 20–30) participated in the study after giving written informed consent. All participants were right-handed (as assessed by an adapted German version the Edinburgh Handedness Inventory; Oldfield, 1971) and had normal or corrected-to-normal vision. One further participant was excluded from the final data analysis due to excessive EEG artifacts.

Materials

In order to ensure maximal comparability with Keller and Sorace's (2003) behavioural findings, we used exactly the same verbs as in their study, namely eight from each of the verb classes in Table 1. For each single verb, we constructed 10 different sentence contexts of the form NP/AUX/ADVERB/PAST PARTICIPLE, thus resulting in 80 sentence frames per condition. Each of these frames was presented once with *haben* ("to have") and once with *sein* ("to be"). The stimulus material for the present study thus comprised 640 sentences (160 sentences for each verb class; 80 with *haben* and 80 with *sein*). These were divided into two lists such that, on each list, each critical sentence frame was presented only once (either with *haben* or *sein*). Each list was supplemented by 80 filler sentences, 40 of which were semantically unacceptable. Every list (containing a total of 400 sentences) was presented in four different constrained randomisations, thus resulting in eight different versions counterbalanced across participants. The randomisation constraints were as follows: trials from the same condition were separated by at least two intervening trials of other types; repeated occurrences of single verbs were separated by at least five intervening trials; auxiliaries were repeated in no more than two successive trials; expected yes/no responses for the probe detection task were identical in no more than three successive trials; each block contained an approximately equal number of trials from each condition and an approximately equal number of expected yes and no responses for the probe detection task.

Procedure

Participants sat in a dimly lit, sound-attenuating booth. Stimuli were presented visually in the centre of a computer screen in yellow letters against a blue background.

Each trial began with the presentation of a fixation asterisk [1,000 ms + 200 ms inter-stimulus interval (ISI)]. NPs were presented as phrases (presentation: 450 ms, ISI: 100 ms), all other elements word-by-word. Auxiliaries and adverbs were presented for 400 ms with an ISI of 100 ms, verbs were presented for 450 ms. The presentation of a sentence was followed by 550 ms of blank screen, after which participants were required to complete an acceptability judgment task (signalled by a question-mark) by pressing one of two push buttons for "yes" or "no" (the maximal reaction time was 3,000 ms, after which the task was automatically terminated). After a participant's reaction, there was a pause of 200 ms before a probe word appeared on the screen. Here, the task was to decide whether the presented word was part of the previous sentence (50%) or not (50%). After the button press (after a maximal reaction time of 3,000 ms, the trial was automatically terminated), there was an inter-trial interval of 2,500 ms before the next trial started. We used a probe task in addition to the acceptability judgement in order to ensure that participants would need to read the sentences completely rather than focusing exclusively on the auxiliary-verb combination (a strategy that would have been adequate for the experimental trials, though not the filler trials). Participants were asked to avoid movements and eye-blinks during the presentation of the sentences. The experimental session began with a short training session followed by eight experimental blocks of 50 trials each, between which the participants took short breaks. Each experimental session lasted approximately 2 hours (including electrode preparation).

EEG-recording and preprocessing

The EEG was recorded by means of 64 Ag/AgCl-electrodes fixed at the scalp by means of an elastic cap (Electrocap International, Eaton, OH, USA). The ground electrode was positioned at C2. Recordings were referenced to the left mastoid but re-referenced to linked mastoids offline. The electrooculogram (EOG) was monitored by means of electrodes placed at the outer canthus of each eye for the horizontal EOG and above and below the participant's left eye for the vertical EOG. Electrode impedances were kept below 5 kΩ. All EEG and EOG channels were amplified using a BrainVision BrainAmp amplifier (DC, high cutoff 250 Hz) and recorded with a digitization rate of 500 Hz. In order to exclude slow signal drifts, the raw EEG data were filtered offline with a 0.3–20 Hz band pass. Automatic and manual rejections were carried out to exclude periods containing movement or technical artifacts (the automatic EOG rejection criterion was 40 μV). Trials for which the probe detection task was answered incorrectly were also excluded from further analysis.

Data analysis

For the acceptability judgments, mean acceptability ratings were computed (as for the ERPs, trials with an incorrect response to the probe detection were excluded). For the probe detection task, we analysed mean error rates. In both cases, we computed a repeated measures analysis of variance (ANOVA) containing the condition factors VERB and AUXiliary and the random factors participants (F_1) and items (F_2). We refrained from analyzing reaction times as the behavioural tasks were not directly time-locked to the critical word in the sentences.

Average ERPs were calculated per condition per participant from −200 to 1,000 ms relative to the onset of the critical sentence-final verb, before grand-averages were computed over all participants. For the statistical analysis, repeated measures ANOVAs were computed using the condition factors VERB and AUX and the topographical

factor region of interest (ROI). Lateral ROIs were defined as follows: left-anterior (F7 F5 F3 FT7 FC5 FC3), left-central (T7 C5 C3 TP7 CP5 CP3), left-posterior (P7 P5 P3 PO7 PO3 O1), right-anterior (F8 F6 F4 FT8 FC6 FC4), right-central (T8 C6 C4 TP8 CP6 CP4), and right-posterior (P8 P6 P4 PO8 PO4 O2). For midline electrodes, we used two ROIs: fronto-central (FZ, FCZ, CZ) and centro-parietal (CPZ, PZ, POZ). In order to evaluate the possibility of effects changing over the course of the experiment (e.g. due to the fact that HAVE was correct for one of the rigid verb types while BE was correct for the other two), we additionally included the factor experiment HALF (first vs. second) in the analyses. All statistical analyses were carried out in a hierarchical manner, that is, only significant interactions ($p < .05$) were resolved. To avoid type I errors resulting from violations of sphericity, we report Greenhouse–Geisser corrected probability levels for all effects with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959).

RESULTS

Behavioural data

Mean acceptability ratings for the judgement task and mean error rates for the probe detection task are given in Table 1.

The analysis of the acceptability rates revealed a significant interaction VERB × AUX, $F_1(3, 93) = 246.2$, $p < .001$; $F_2(3, 316) = 59.23$, $p < .001$. Resolving the interaction by VERB showed simple effects of auxiliary type for CH-LOC verbs, $t1(31) = 80.42$, $p < .001$; $t2(79) = 65.39$, $p < .001$; CON-PROC verbs, $t1(31) = -91.06$, $p < .001$; $t2(79) = -79.62$, $p < .001$; and CH-STATE-PRE verbs, $t1(31) = 49.74$, $p < .001$; $t2(79) = 36.12$, $p < .001$. No significant difference was observed for the CH-STATE-UN verbs ($t1/t2 < 1$).

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

ERP data

Grand average ERPs at the position of the critical verb are shown for each verb class in Figures 2–5.

As is apparent from Figures 2, 4, and 5, sentences with a dispreferred auxiliary led to a biphasic N400-late positivity response for CH-LOC, CH-STATE-PRE, and CON-PROC verbs. By contrast, there was no observable difference for CH-STATE-UN (Figure 3). Statistical analyses were conducted in two time windows: 380–530 ms for the N400 and 750–900 for the late positivity. (Note that, since this was the first ERP study on auxiliary selection and, therefore, no a priori information was available for time window definition, time windows were selected via visual inspection.) The results of the statistical analyses are summarised in Tables 2 and 3.

The analysis in Tables 2 and 3 shows that auxiliary selection violations for the three rigid verb types (CH-LOC, CH-STATE-PRE, and CON-PROC) engendered a broadly distributed negativity with a posterior focus (N400) followed by a late positivity with a right-posterior focus. While the analyses for the N400 time window did not provide any compelling evidence for differing topographical distributions across the different verb classes, the late positivity was less widely distributed for CH-STATE-PRE verbs than for CH-LOC and CON-PROC verbs, since it was confined to posterior ROIs for

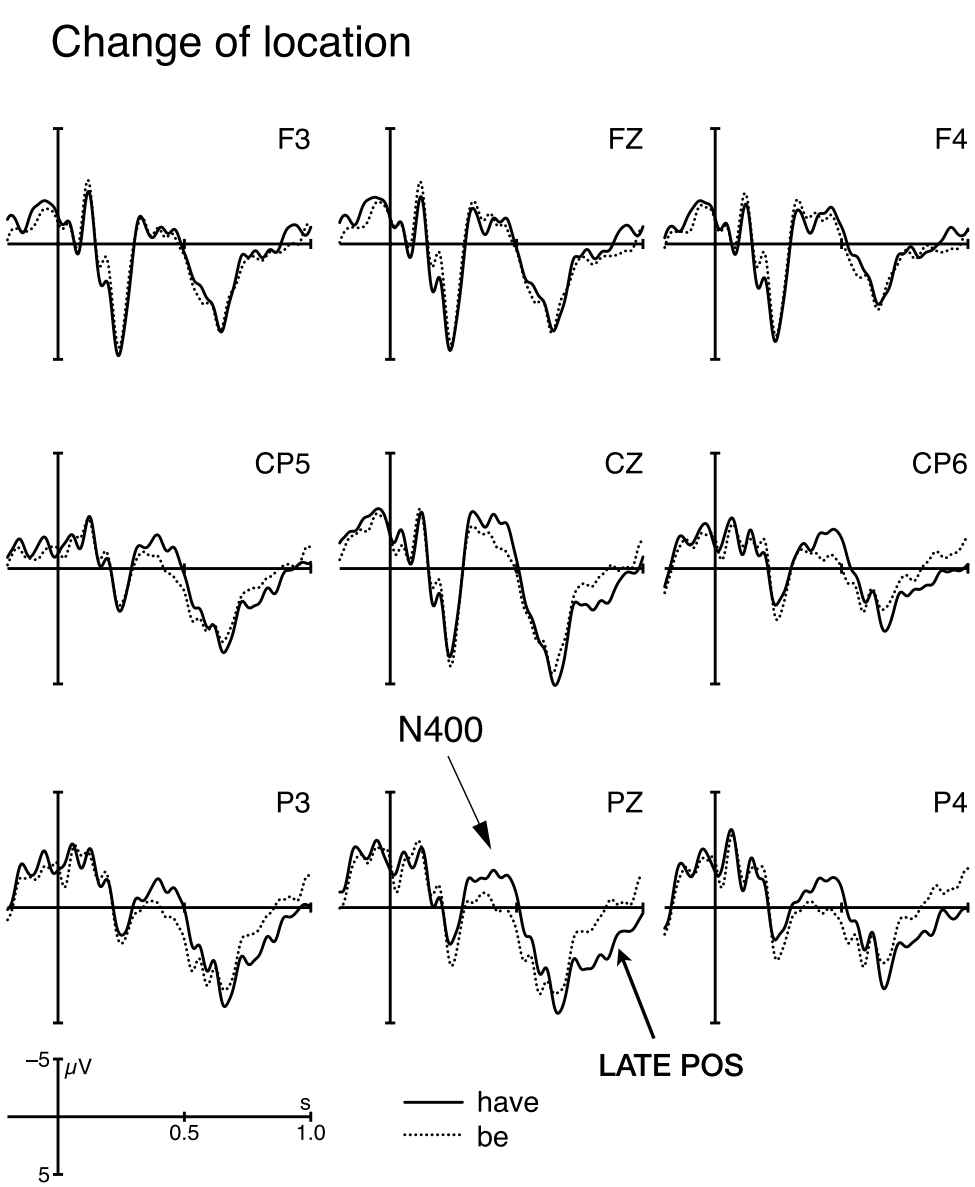

Figure 2. Grand average ERPs at the position of the critical verb (onset at the vertical bar) for CH-LOC verbs in sentences with HAVE vs. BE. Negativity is plotted upwards.

CH-STATE-PRE verbs but also reached significance in central regions for the other two verb types.

In the N400 time window, no interactions with the factor HALF reached significance, thus indicating that the N400 was stable across the course of the experiment. For the late positivity, interactions with HALF were observed. However, these were not due to a qualitative change of effects between the two halves of the experiment, but rather resulted from the fact that the positivity was topographically more restricted (to right centro-parietal sites) in the second half of the study. Crucially, a significant positivity effect was observed for all three rigid verb types in both halves,

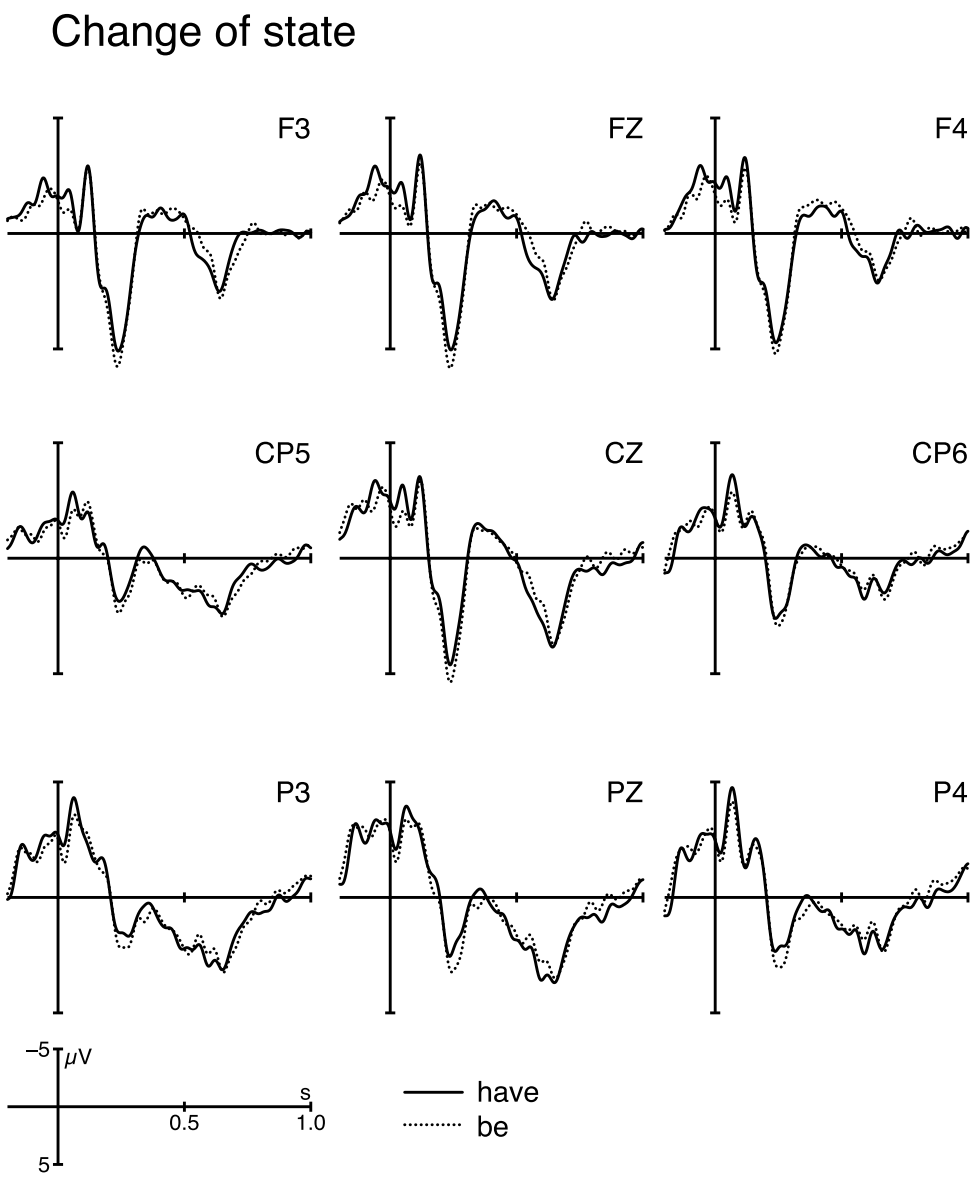

Figure 3. Grand average ERPs at the position of the critical verb (onset at the vertical bar) for CH-STATE-UN verbs in sentences with HAVE vs. BE. Negativity is plotted upwards.

thus indicating that, qualitatively, the pattern was similar across the course of the experiment, though the positivity was somewhat stronger in the first half.

Single trial analyses: Correlations between the ERP findings and the behavioural data

The CH-STATE-UN verbs showed an apparent indeterminacy with respect to auxiliary choice both in terms of acceptability ratings (which had mean values between 60 and 65% for both HAVE and BE across participants and items) and ERP responses (which showed no difference between HAVE and BE in the grand averages).

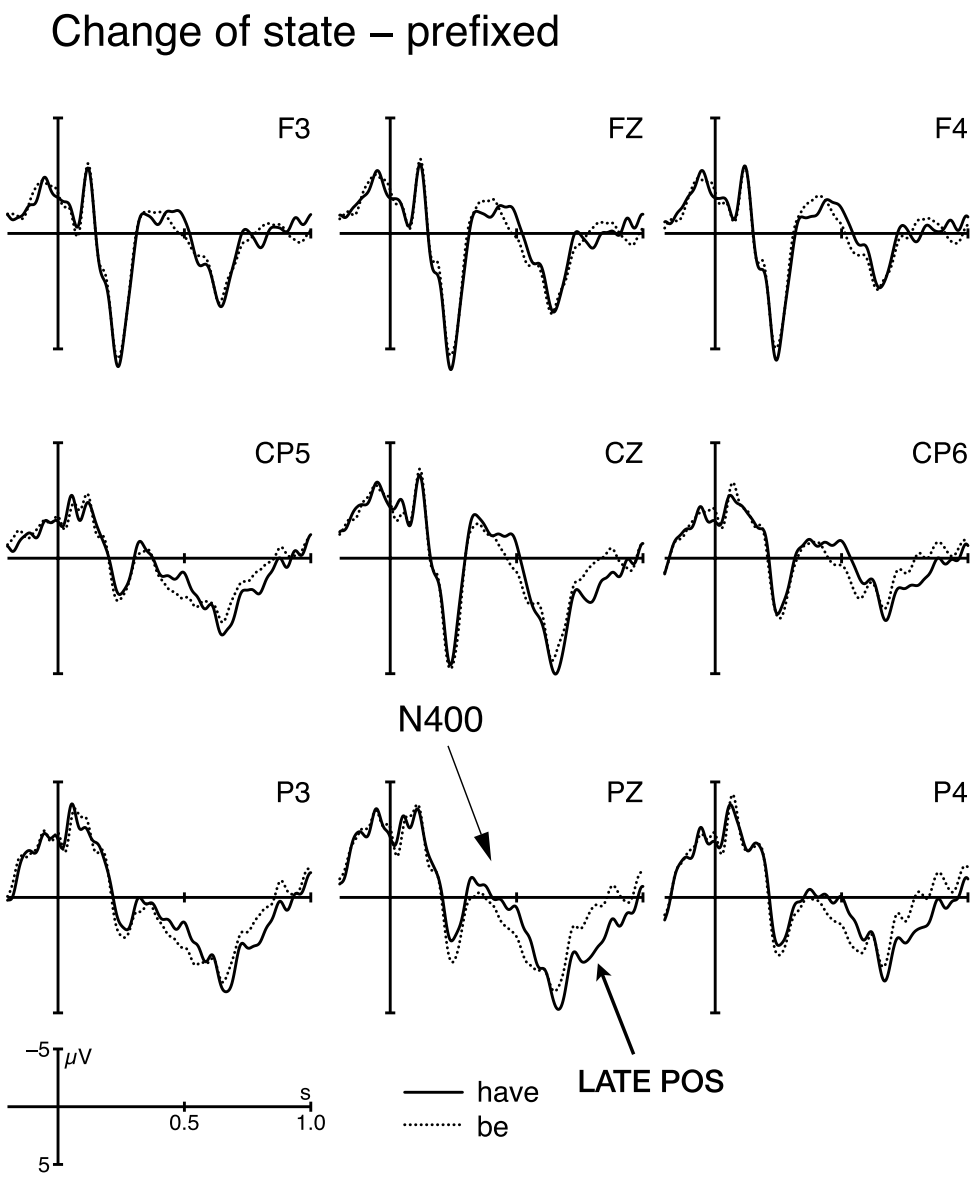

Figure 4. Grand average ERPs at the position of the critical verb (onset at the vertical bar) for CH-STATE-PRE verbs in sentences with HAVE vs. BE. Negativity is plotted upwards.

In this context, there appear to be two possible scenarios for why the grand average ERPs did not show a significant effect of AUX: (1) as predicted by the Indeterminacy Hypothesis, this verb class is truly underspecified such that there is no conflict with the expectations set up by either HAVE or BE; or (2) the absence of an effect in the grand averages is a result of the averaging procedure, that is, a product of averaging over gradient responses which differ on a trial-to-trial basis. The second scenario predicts that effects of AUX should be revealed by a more fine-grained analysis of the ERP data for the CH-STATE-UN verbs.

In order to test whether this is indeed the case, we analysed the ERP results for this verb class using linear mixed-effects models (Baayen, 2008; Baayen, Davidson, &

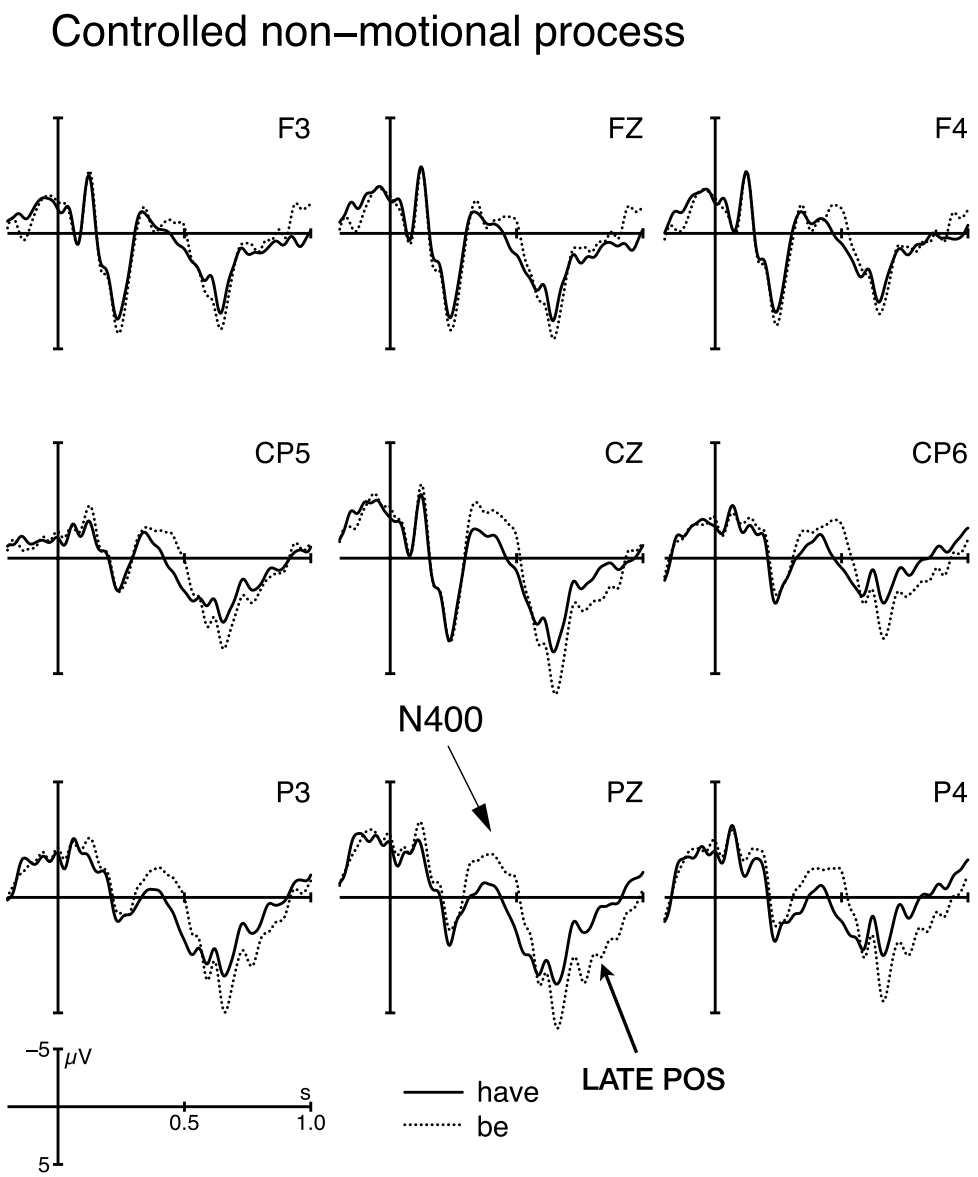

Figure 5. Grand average ERPs at the position of the critical verb (onset at the vertical bar) for CON-PROC verbs in sentences with HAVE vs. BE. Negativity is plotted upwards.

Bates, 2008). Models included the fixed factor AUX and the crossed random factors participants and items. Furthermore, we included by-participant acceptability (i.e. individual participants' ratings of CH-STATE-UN verbs averaged across all sentence frames) and by-item acceptability ratings (i.e. acceptability ratings per auxiliary type for each individual sentence frame with a CH-STATE-UN verb averaged across all participants) as predictors. The analysis was restricted to posterior electrode sites (i.e. the P, PO and O lines) in order to (1) avoid the problem of having to use a particular ROI as reference level if a fixed factor ROI were included, and (2) increase interpretability of the results by avoiding a large number of higher-order interactions. This restriction is in accordance with our prediction of N400 and late positivity effects

TABLE 2
Summary of the statistical analyses for the ERP data in time window 1 (380–530 ms post onset of the critical sentence-final verb). In order to increase readability, only significant effects are reported (i.e. if an effect is not listed, it did not reach significance)

Midline electrodes	*Lateral electrodes*
WIN 1	
VERB: $F(3, 93) = 3.45$*	VERB: $F(3, 93) = 6.80$***
VERB × ROI: $F(3, 93) = 40.09$***	VERB × ROI: $F(15, 465) = 29.45$***
VERB × AUX: $F(3, 93) = 10.43$***	VERB × AUX: $F(3, 93) = 16.80$***
VERB × AUX × ROI: $F(3, 93) = 40.09$***	
Fronto-central VERB × AUX: $F(3, 93) = 8.11$***	CH-LOC AUX: $F(1, 31) = 20.60$***
CH-STATE-PRE AUX: $F(1, 31) = 5.62$*	CH-STATE-PRE AUX: $F(1, 31) = 5.33$*
CON-PROC AUX: $F(1, 31) = 13.46$***	CON-PROC AUX: $F(1, 31) = 21.69$***
Centro-parietal VERB × AUX: $F(3, 93) = 10.75$***	
CH-LOC AUX: $F(1, 31) = 11.87$**	
CH-STATE-PRE AUX: $F(1, 31) = 5.39$*	
CON-PROC AUX: $F(1, 31) = 13.52$***	

***$p < .001$; **$p < .01$; *$p < .05$. Greenhouse–Geisser corrected p-values are reported for all effects with more than one degree of freedom in the numerator.

(both of which show a posterior maximum) and with the topographical distribution of the effects in the rigid verb conditions (see Tables 2 and 3).

Model fitting proceeded as follows. For each time window, we first fit a base model including only AUX. We then proceeded to add by-item-acceptability and by-subject-acceptability in separate models and, using likelihood-ratio tests, assessed for each whether it improved the fit over the base model. If this was the case, we proceeded to add the second acceptability predictor to each of the new models and again assessed whether this improved the fit. For this assessment of nested models relative to one another, variables of interest and all higher-order interactions were included in each of the models. Once the best-fitting model had been determined via these nested comparisons, we proceeded to apply model simplification (i.e. the removal of nonsignificant higher-order interactions) until the minimal adequate model was reached. Analyses were conducted using R (R Development Core Team, 2007) and the lme4 package for linear mixed-effects models (Bates, 2005). Uncertainty in the parameter estimation was assessed using simulations that were computed by means of the arm package (Gelman et al., 2009).

Time window 1

In the first time window, the inclusion of by-item acceptability led to a significant improvement of the fit over the base model, $\chi^2(2) = 65.96$, $p < .001$, while the inclusion of by-participant acceptability did not, $\chi^2(2) = 0.81$, $p > .6$. Additionally including by-participant acceptability further improved the fit over that of the model that only included by-item acceptability and AUX, $\chi^2(4) = 56.76$, $p < .001$. Since the best-fitting model involved significant interactions with the factor AUX, we also conducted separate analyses for HAVE and BE. These revealed that the interactions were primarily driven by HAVE. Model parameters are summarised in Table 4 and visualised in Figure 6. As is apparent from the model and the figure, the interaction of by-participant and by-item acceptability predicts lower (i.e. more positive) N400 amplitudes for HAVE only. In order to examine the underlying cause of the interaction, which suggests that the effect of by-item acceptability on ERP amplitude

TABLE 3
Summary of the statistical analyses for the ERP data in time window 2 (750–900 ms post onset of the critical sentence-final verb)

Midline electrodes	*Lateral electrodes*
WIN 2	
VERB: $F(3, 93) = 8.03$***	VERB: $F(3, 93) = 6.93$**
AUX: $F(3, 93) = 5.47$*	VERB × ROI: $F(15, 465) = 2.93$*
VERB × ROI: $F(3, 93) = 3.18$*	VERB × AUX: $F(3, 93) = 5.08$**
VERB × AUX: $F(3, 93) = 9.15$***	VERB × AUX × HALF: $F(3, 93) = 3.56$*
VERB × AUX × ROI: $F(3, 93) = 44.28$***	VERB × AUX × ROI: $F(15, 465) = 14.20$***
Centro-parietal: VERB × AUX: $F(3, 93) = 28.75$***	VERB × AUX × HALF × ROI: $F(15, 465) = 2.69$*
CH-LOC AUX: $F(1, 31) = 20.11$***	
CH-STATE-PRE AUX: $F(1, 31) = 27.83$***	Half 1: VERB × AUX × ROI: $F(15, 465) = 14.71$***
CON-PROC AUX: $F(1, 31) = 49.73$***	Left-central VERB × AUX: $F(3, 93) = 4.20$*
VERB × AUX × HALF: $F(3, 93) = 2.95$*	CH-LOC AUX: $F(1, 31) = 13.00$**
Half 1: VERB × AUX: $F(3, 93) = 10.88$***	Right-central VERB × AUX: $F(3, 93) = 11.99$***
CH-LOC AUX: $F(1, 31) = 8.34$**	CH-STATE-PRE AUX: $F(1, 31) = 9.92$**
CH-STATE-PRE AUX: $F(1, 31) = 6.00$*	CON-PROC AUX: $F(1, 31) = 19.93$***
CON-PROC AUX: $F(1, 31) = 15.70$***	Left-posterior VERB × AUX: $F(3, 93) = 14.32$***
Half 2: VERB × AUX: $F(3, 93) = 3.81$*	CH-LOC AUX: $F(1, 31) = 16.46$**
CH-STATE-PRE AUX: $F(1, 31) = 8.14$**	CH-STATE-PRE AUX: $F(1, 31) = 4.61$*
	CON-PROC AUX: $F(1, 31) = 15.04$***
	Right-posterior VERB × AUX: $F(3, 93) = 22.25$***
	CH-LOC AUX: $F(1, 31) = 15.05$**
	CH-STATE-PRE AUX: $F(1, 31) = 6.63$*
	CON-PROC AUX: $F(1, 31) = 31.39$***
	Half 2: VERB × AUX × ROI: $F(15, 465) = 5.48$**
	Left-anterior VERB × AUX: $F(3, 93) = 4.40$*
	CON-PROC AUX: $F(1, 31) = 11.08$**
	Right-central VERB × AUX: $F(3, 93) = 6.36$**
	CH-STATE-PRE AUX: $F(1, 31) = 9.92$**
	CON-PROC AUX: $F(1, 31) = 6.34$*
	Right-posterior VERB × AUX: $F(3, 93) = 6.62$**
	CH-LOC AUX: $F(1, 31) = 7.33$*
	CH-STATE-PRE AUX: $F(1, 31) = 9.36$*
	CON-PROC AUX: $F(1, 31) = 6.45$*

***$p < .001$; **$p < .01$; *$p < .05$. Greenhouse–Geisser corrected p-values are reported for all effects with more than one degree of freedom in the numerator.

is dependent on by-participant acceptability, we performed a median split on by-participant acceptability and then fit separate models including by-item acceptability for the two groups of participants (i.e. the "high-acceptability group", those judging CH-STATE-UN with HAVE with above-median acceptability, and the "low-acceptability group", those judging CH-STATE-UN with HAVE with below-median acceptability).[7] These revealed an effect of by-item acceptability only for the high-acceptability group (model estimates: intercept 0.34; 95% CI: −0.70 to 1.44, $p > .5$; by-item acceptability 1.95; 95% CI: 0.71–3.02; $p < .01$), but not for the

[7]Note that the median splits provide a useful means of examining the underlying cause of the interactions. They should not, however, be taken to suggest that the participants or items fall into two distinct classes ("low" vs. "high" acceptability). The advantage of using linear effects modelling is the ability to model the effect of continuous variables directly rather than having to group their effects into discrete classes (Baayen et al., 2008).

TABLE 4
Parameter values for fixed effects in the best-fitting linear mixed-effects models of mean amplitude values for the CH-STATE-UN verbs in the N400 time window (380–530 ms) and associated confidence intervals and *p*-values. Confidence intervals and *p*-values were obtained by MCMC sampling (Baayen et al., 2008). Note that only data from the posterior electrodes entered this model (see the main text) and that, for the factor auxiliary type, BE was the reference level. For reasons of readability, only significant effects are reported (note that the separate analysis for BE revealed no significant effects)

	Estimate	*95% CI-lower*	*95% CI-upper*	*p-Value*
Global analysis				
AUX-(have)	2.85	1.62	4.13	<.001
AUX-(have):ITM-ACC	−4.03	−5.74	−2.31	<.001
AUX-(have):SUBJ-ACC	−3.85	−5.69	−2.00	<.001
AUX-(have):ITM-ACC:SUBJ-ACC	5.72	3.26	8.21	<.001
HAVE				
ITM-ACC	−2.88	−4.22	−1.40	<.001
SUBJ-ACC	−6.14	−8.93	−3.21	<.001
SUBJ-ACC:ITM-ACC	6.75	5.04	8.42	<.001

SUBJ-ACC, participant-based acceptability; ITM-ACC, item-based acceptability; AUX, auxiliary type.

low-acceptability group (model estimates: intercept 1.73; 95% CI: 0.64–2.80, $p < .01$; by-item acceptability 0.19; 95% CI: −1.11 to 1.42; $p > .7$). Thus, the reduction of N400 amplitude with increasing by-item acceptability is driven primarily by those participants who have a higher tendency to accept CH-STATE-UN verbs with HAVE.

Time window 2

In the second time window, both the model including by-item acceptability and that including by-participant acceptability led to a significant improvement of fit over the base model (model including by-item acceptability vs. base model: $\chi^2(2) = 58.79$, $p < .001$; model including by-participant acceptability vs. base model: $\chi^2(2) = 19.65$, $p < .001$). However, the additional inclusion of the second acceptability predictor further improved both of these models (model including both by-item and by-participant acceptability vs. model including only by-item acceptability: $\chi^2(4) = 49.27$, $p < .001$; model including both by-item and by-participant acceptability vs. model including only by-participant acceptability: $\chi^2(4) = 88.41$, $p < .001$). In the full model, the three-way interaction (AUX × SUBJ-ACC × ITM-ACC) did not reach significance and was thus removed in the process of model simplification. The minimal adequate model thus resulting is summarised in Table 5 and visualised in Figure 7. Since this model showed an interaction between AUX and by-participant acceptability, we again fit separate models for HAVE and BE, including only by-participant acceptability. As in the earlier time window, effects were driven by HAVE, as is visualised in Figure 8. This effect was due to larger amplitudes of the late positivity with increasing by-participant acceptability.

As for time window 1, we assessed the cause of the by-participant and by-item acceptability interaction in the global analysis for time window 2 by performing a median split on by-participant acceptability (in this case, over both auxiliaries, since there was no three-way interaction with AUX) and fitting separate models including by-item acceptability for the high- and low-acceptability groups. Only by-item acceptability was included in the models (i.e. not AUX or the interaction between

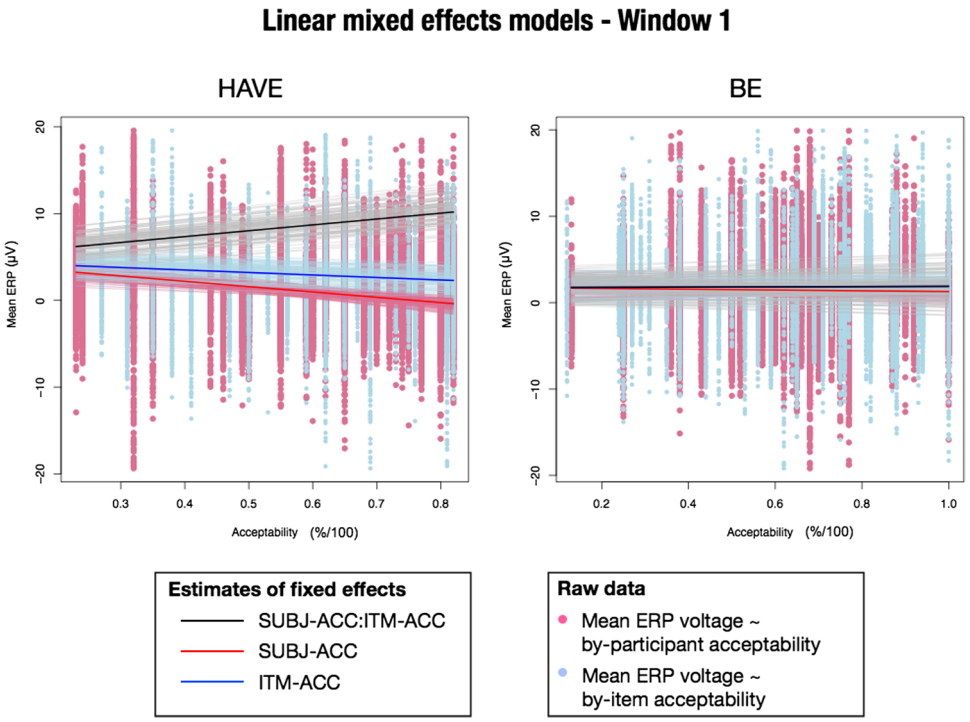

Figure 6. Illustration of the best-fitting linear mixed-effects models for HAVE (left panel) and BE (right panel) in the N400 time window (see also Table 4). Points designate raw data points (light red: mean ERP voltages ~ by-participant acceptability; light blue: mean ERP voltages ~ by-item acceptability) and lines illustrate the model fits (red: by-participant acceptability; blue: by-item acceptability; black by-participant acceptability: by-item acceptability). Shading around the regression lines indicates the regression uncertainty as measured via 100 random model simulations produced by the sim-function (Gelman et al., 2009). Note that, for BE, modelling revealed no significant effects.

the two factors) because of the absence of the three-way interaction in the global analysis. These separate analyses revealed effects of by-item acceptability for both groups, though the effect was stronger in the high-acceptability group (model estimates: intercept 1.49; 95% CI: 0.81 to 2.19, $p < .0001$; by-item acceptability -1.91; 95% CI: -2.28 to -1.54; $p < .0001$) than in the low-acceptability group (model estimates: intercept 0.38; 95% CI: -0.17 to 0.91, $p > .1$; by-item acceptability -0.50; 95% CI: -0.86 to -0.14; $p < .01$). In both groups, higher acceptability ratings correlated with smaller positivity amplitudes.

Taken together, the global model and the model for HAVE in this time window suggest that, overall, the interaction of by-participant and by-item acceptability correlates with lower (i.e. more negative) late positivity amplitudes. More precisely, higher by-item acceptability correlates with smaller late positivity amplitudes and this correlation is more pronounced for participants with higher acceptability ratings for the CHANGE-STATE-UN verbs. For HAVE only, by contrast, increasing by-participant acceptability correlates with higher late positivity amplitudes.

One potential concern with the analyses presented above is that the model parameter estimates—and particularly the opposite directionality of the main effects and interactions—may have resulted from a high collinearity of the main effects of acceptability and the interaction between the two acceptability variables and the resulting inability of the model to accurately predict the correlated predictors. In order

TABLE 5
Parameter values for fixed effects in the best-fitting linear mixed-effects models of mean amplitude values for the CH-STATE-UN verbs in the late positivity time window (750–900 ms) and associated confidence intervals and *p*-values. Confidence intervals and *p*-values were obtained by MCMC sampling (Baayen et al., 2008). Note that only data from the posterior electrodes entered this model (see the main text) and that, for the factor auxiliary type, BE was the reference level. For reasons of readability, only (at least marginally) significant effects are reported (note that the separate analysis for BE revealed no significant effects)

	Estimate	*95% CI-lower*	*95% CI-upper*	p-*Value*
Global analysis				
Intercept	−0.77	−1.61	0.03	<.07
SUBJ-ACC	2.02	1.03	3.06	<.001
ITM-ACC	1.35	0.42	2.23	<.01
AUX-(have):SUBJ-ACC	1.52	0.54	2.45	<.001
SUBJ-ACC:ITM-ACC	−3.28	−4.51	−2.09	<.001
HAVE				
Intercept	−1.54	−3.14	0.11	<.07
SUBJ − ACC	3.06	0.49	5.70	<.02

SUBJ-ACC, participant-based acceptability; ITM-ACC, item-based acceptability; AUX, auxiliary type.

to assess this possibility, we recalculated the analyses with centred acceptability measures (i.e. with by-participant and by-item acceptabilities with the mean subtracted from each) to reduce the collinearity. These additional analyses, which are presented in detail in Appendix 1, revealed very similar results to the analyses presented above.

DISCUSSION

The present ERP study used split-intransitivity in order to study the processing of flexibility at the syntax–semantics interface during online language comprehension. For sentences including a dispreferred auxiliary, we observed a biphasic N400-late positivity pattern. This pattern was engendered by core unergative verbs (verbs of controlled nonmotional process, CON-PROC) and core unaccusative verbs (verbs of change of location, CH-LOC), that is, verbs with an inherent lexical specification of the key semantic feature of *telicity*, and for lexically indeterminate verbs with a telicity-inducing prefix (prefixed change of state verbs, CH-STATE-PRE). Finally, sentences with lexically indeterminate verbs (unprefixed change of state verbs, CH-STATE-UN) did not show any differences between BE and HAVE in grand average ERPs. These verbs also led to gradient acceptability ratings (between approximately 60 and 65% for both HAVE and BE in mean ratings across participants and items). An additional analysis of the ERP data for the CH-STATE-UN verbs using linear mixed-effects models revealed an interaction between auxiliary selection and acceptability ratings. Importantly, the nature of the interaction differed between the two time windows. In the N400 window, all effects were driven by HAVE, and N400 amplitudes were lower with increasing by-item acceptability with HAVE, particularly for participants who tended to accept CH-STATE-UN verbs with HAVE. In the late positivity window, the interaction of by-participant and by-item acceptability predicted reduced late positivity amplitudes—now across both auxiliary types. Again, the ERP amplitude was reduced with increased by-item acceptability and particularly so for participants who judged sentences with CH-STATE-UN verbs with a higher-than median

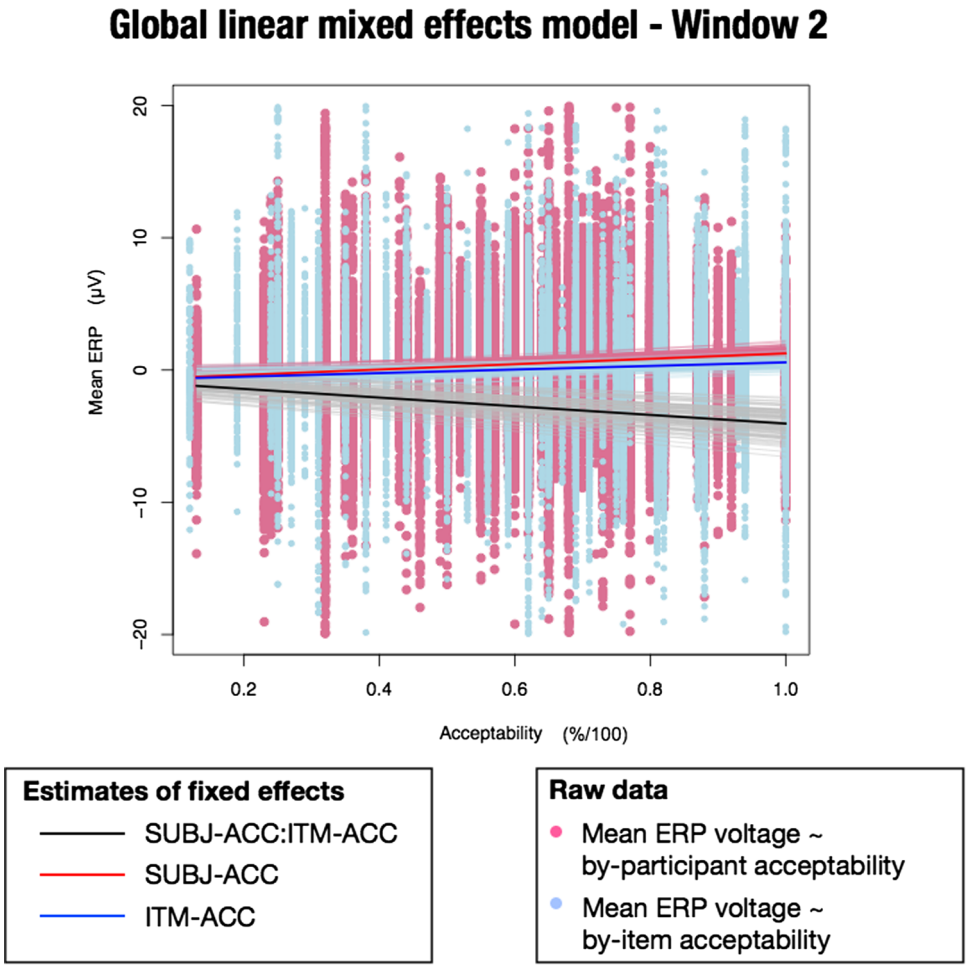

Figure 7. Illustration of the best-fitting global linear mixed-effects model in the late positivity time window (see also Table 5). Points designate raw data points (light red: mean ERP voltages ~ by-participant acceptability; light blue: mean ERP voltages ~ by-item acceptability) and lines illustrate the model fits (red: by-participant acceptability; blue: by-item acceptability; black by-participant acceptability: by-item acceptability). Shading around the regression lines indicates the regression uncertainty as measured via 100 random model simulations produced by the sim-function (Gelman et al., 2009).

acceptability. In addition, however, separate analyses for the two auxiliaries in the late time window (motivated by an AUX × SUBJ-ACC interaction) revealed that, for HAVE, by-participant acceptability correlated positively with ERP amplitude (i.e. higher acceptability correlated with larger positivity effects).

In the following, we first discuss the overall ERP pattern before turning more specifically to the single trial-based analyses of the flexible verbs and their potential functional significance.

Flexibility versus rigidity in the form-to-meaning mapping: No evidence for qualitative differences

The grand average ERPs did not provide any evidence for a qualitative distinction in the processing of flexible versus rigid syntax–semantics mappings. They also indicate

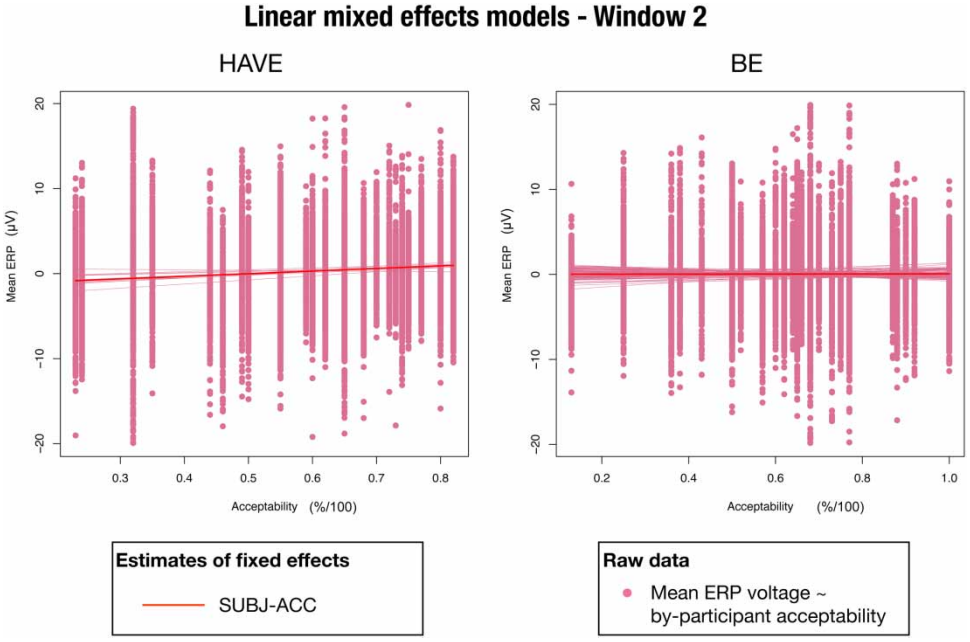

Figure 8. Illustration of the best-fitting linear mixed-effects models for HAVE (left panel) and BE (right panel) in the late time window (see also Table 5). Points designate raw data points (mean ERP voltages ~ by-participant acceptability) and lines illustrate the model fits (by-participant acceptability). Note that only by-participant acceptability is included as a predictor here since this was the only predictor to interact with auxiliary type in the global model for this time window (see Table 5). Note that, for BE, modelling revealed no significant effects.

that lexical and compositional specifications of telicity have a qualitatively similar effect on the neurophysiological responses observed here, since auxiliary selection violations engendered a qualitatively similar biphasic N400-late positivity pattern for all categorical verbs irrespective of whether they were lexically specified core verbs (CH-LOC; CON-PROC) or lexically indeterminate verbs for which the choice of auxiliary was determined by the addition of a telicity-inducing prefix (CH-STATE-PRE). This result can—at least tentatively—be viewed as converging support for a linking-based interpretation of the N400 in auxiliary-selection violations as opposed to an interpretation based purely on lexical preactivation. Assuming that complex morphological forms are decomposed (for a recent review of the evidence for decomposition in the processing of derivational morphology, see Bornkessel-Schlesewsky & Schlesewsky, 2009a), the absence of lexical preactivation for the stem of CH-STATE-PRE verbs should have manifested itself in the electrophysiological processing signature for these verbs.[8] However, the CH-STATE-PRE class showed a qualitatively similar pattern to the (lexically specified) core verbs examined in the present study. This suggests that the N400 observed for auxiliary selection mismatches does not reflect—or is least not exclusively an index of—the relative ease or difficulty

[8] We do not think that this result can be explained in terms of prefix preactivation, since the prefixes used in our materials (*ver-*, *an-*, *auf-*, and *er-*) not only mark telicity, but also serve other functions and can thus occur with atelic verbs [e.g. *verspielen*, 'to lose through gamling/to give something away', *verlaufen*, "to get lost", *ansagen* ("to announce"), *anmalen* ("to paint something")]. Thus, it is the combination of prefix and verb that results in the telic reading, rather than the prefix in and of itself.

of lexical access for a particular verb class in the context of HAVE or BE (see Lau et al., 2008, for an account of the N400 based on lexical preactivation), but that it rather correlates with compositional processes in the linking between form and meaning (for linking-related N400 effects in language comprehension, see e.g. Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2008a, 2009b). In other words, we assume that the N400 effect for core and prefixed verbs is most parsimoniously explained in terms of the anticipation of particular verb classes (i.e. linking classes with particular aspectual properties) rather than the preactivation of individual words (for evidence of N400 modulations linked to the anticipation of particular verb classes, see Bornkessel-Schlesewsky & Schlesewsky, 2008b).[9] Furthermore, and in line with previous results, we assume that the late positivity reflects a categorisation process by means of which the sentences with a dispreferred auxiliary are classified as ill-formed (Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky et al., 2011; Kretzschmar, 2010).

An anonymous reviewer suggests an alternative interpretation of the late positivity in terms of reanalysis processes. Late positive ERP effects (P600s) have long been associated with the resolution of garden path effects (e.g. Osterhout & Holcomb, 1992) and, accordingly, been viewed as reflecting processes of reanalysis and/or repair (for reviews, see, for example, Friederici, 2002; Kutas, Van Petten, & Kluender, 2006). From this perspective, it appears reasonable to posit that the late positivity effects observed here could reflect the reanalysis processes that are required when the aspectual ambiguity (telic vs. nontelic event) set up by the auxiliary is resolved towards the dispreferred reading at the position of the verb. However, since reanalysis-based accounts of the late positivity have typically focused on *syntactic* reanalysis, it is not entirely clear what would be reanalysed, since even under the assumption of distinct syntactic configurations with BE and HAVE, the structure would be fixed once the auxiliary is reached. We also favour the well-formedness categorisation account for several additional reasons. Firstly, there is no one-to-one mapping between reanalysis processes and the late positivity, since—at least certain types of—reanalyses have also been shown to correlate with other ERP responses (most notably N400 effects) in some studies (e.g. Bornkessel et al., 2004; Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008). Secondly, a categorisation-based account appears broader in scope since it can also derive the observation of positivity effects in unambiguous sentences, as documented, for example, in the literature on "semantic P600" effects (for reviews, see for example, Bornkessel-Schlesewsky & Schlesewsky, 2008a; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009). Finally, an account based on well-formedness is highly compatible with the observation of higher late positivity amplitudes for less acceptable sentences in the trial-by-trial analysis of the CH-STATE-UN conditions: recall that we observed weaker positivity responses with increasing by-item acceptability and that this correlation was stronger for participants who showed a higher general acceptability for the sentence type in question. In spite of these arguments in favour of a well-formedness-based account, however, we cannot conclusively rule out a reanalysis-based view, since the present study was not designed to tease apart these competing interpretations of the late positivity.

[9] In this way, the N400 effects observed here go beyond the typical predictability effects reported in the N400 literature, in which the prediction of individual words is crucial (e.g. DeLong, Urbach, & Kutas, 2005; Otten, Nieuwland, & van Berkum, 2007; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004).

Indeterminacy or enrichment in the processing of flexible verbs? Single trial-based correlations between neurophysiological responses and acceptability

Perhaps the most interesting finding of the present study was that, for the verbs which did not show a preference for either HAVE or BE in terms of acceptability ratings (CH-STATE-UN) and also did not engender auxiliary-selection-based differences in grand average ERPs, ERP amplitudes in the N400 and late positivity time windows showed an interaction between auxiliary choice and individual by-item/by-subject acceptability in a linear mixed model analysis. This finding suggests that the absence of an effect in the average ERP responses was likely due to the averaging procedure rather than resulting from the processing system's general indifference to auxiliary selection with this particular class of verbs. Regarding the mechanisms resulting in gradient behaviour for CH-STATE-UN verbs, this finding supports the Enrichment Hypothesis rather than the Indeterminacy Hypothesis. Indeed, results were in line with the prediction of this hypothesis that a possible asymmetry between auxiliaries should result from effects being driven by HAVE rather than BE, since the change of state in CH-STATE-UN verbs means that they imply telicity at some level and are thus closer to the "BE-end" of the SIH. In both time windows, all interactions with auxiliary type were exclusively driven by HAVE.

A closer examination of the results from the single-trial analyses can serve to further illuminate the functional significance of the ERP effects observed here. These analyses revealed three basic results: a correlation between higher acceptabilities (dependent upon both participant- and item-based acceptability) and reduced N400 amplitudes for HAVE; a correlation between higher acceptabilities (dependent upon both participant- and item-based acceptability) and reduced late positivity amplitudes across both auxiliary types; and a correlation between higher participant-based acceptabilities and *increased* late positivity amplitudes.

For the N400, the observation of decreased amplitudes with increasing by-item acceptability (for participants with a higher general tendency to accept CH-STATE-UN verbs with HAVE) provides converging support for the assumption that the N400 effects in the present study index a mismatch in auxiliary-induced expectation for aspectual properties of the verb. Thus, the N400 modulation for the gradient verbs can be explained in a similar manner to those for the core and prefixed verbs, with the added assumption that individual gradient verbs vary with regard to how strongly they are specified for a telic reading and, thereby, the degree of mismatch induced.[10]

[10] Note that the negative correlation between N400 amplitude and acceptability (i.e. the fact that lower N400 amplitudes correlated with higher acceptability ratings) renders an alternative interpretation in terms of polysemy unlikely. At a first glance, higher N400 amplitudes could be viewed as indexing sense selection or similar processes for polysemous gradient verbs. However, from this perspective we should have observed a positive correlation between N400 amplitude and acceptability, since a successful switch to an alternative sense should correlate with a higher likelihood for an "acceptable" rating for that particular trial. A second argument against a polysemy-based explanation stems from the observation that ERP responses to semantically ambiguous as opposed to unambiguous words are qualitatively different from typical N400 effects, showing a more frontal and more lateralised distribution (Hagoort & Brown, 1994). A similar effect has also been reported for ambiguities with regard to the thematic structure of particular verb classes (Bornkessel, 2002), thus further suggesting that the present N400 modulations are not related to semantic ambiguity or sense selection. Finally, the explanation advanced here, which posits a unified explanation for the N400 effect for all verb types examined here, appears preferable to separable explanations (verb class anticipation versus polysemy) on grounds of parsimony.

Strikingly, the late positivity time window showed two distinct patterns with regard to the relationship between acceptability ratings and ERP effects: a negative correlation between acceptability and amplitude of the late positivity across both auxiliaries and, for HAVE only, a negative correlation between by-participant acceptability and amplitude of the late positivity (i.e. larger positivity effects for *more* acceptable sentences). Whereas the former fits well with a well-formedness-based interpretation of the late positivity (e.g. in terms of well-formedness categorisation as suggested by Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky et al., 2011; Kretzschmar, 2010 or in terms of conflict monitoring as argued, for example, by Kolk et al., 2003, and van de Meerendonk et al., 2010),[11] the latter is highly suggestive of a coercion-based interpretation (recall from the introduction that the successful application of coercion should result in higher rather than lower acceptability on a single trial basis). This pattern of results thus provides compelling evidence in favour of two qualitatively distinct late positive effects (see Bornkessel & Schlesewsky, 2006, for the prediction that two such effects should exist, though there was no clear-cut empirical evidence for this at the time). The first of these effects appears to reflect the categorisation of the linguistic stimuli as ill-formed, hence correlating negatively with (the interaction of) by-participant and by-item acceptability of a particular stimulus. The other reflects the process of coercing a telic verb into an activity reading in order to render it compatible with the auxiliary HAVE, hence correlating positively with stimulus acceptability. The fact that this correlation only involved by-participant acceptability of HAVE with CH-STATE-UN verbs suggests that coercion may be more readily available to some individuals than to others. Though this is clearly a speculative suggestion at present, it is in line with the observation that item-based effects on the N400 were modulated by by-participant acceptability ratings, with only those participants who showed a higher general acceptability for CH-STATE-UN verbs with HAVE showing the correlation. A second potentially converging result was reported by Nieuwland et al. (2010), who found that ERP correlates of processing underinformative versus informative scalar statements vary with interindividual differences in pragmatic abilities. These observations suggest two possible alternative scenarios for the positive correlation between late positivity amplitude and by-participant acceptability of CH-STATE-UN verbs with HAVE (though there may well be more): (1) coercion is more readily available to those participants who do not experience a strong mismatch between these verbs and HAVE; and (2) coercion may be more readily available to some individuals in general. Disentangling these possibilities will clearly require further research. However, it seems to us that interindividual variation in pragmatic processing will likely prove relevant in some way, since alternative explanation (1) does not explain why the late positivity did not show a similar dependence on by-participant *and by-item* acceptability for CH-STATE-UN verbs with HAVE as the N400.

In view of previous electrophysiological findings on enriched composition, it is not entirely clear why the coercion process that we assume (shifting of an—albeit weakly specified—telic verb to an activity reading) should correlate with a late positivity rather than an N400. One possible explanation is that, like reference transfer and unlike complement coercion, the shift in lexical aspect required here does not require a type change (as in the case of an entity being shifted to an event; see Pylkkänen & McElree, 2006, for detailed discussion). This assumption also fits well with Brennan and Pylkkänen's (2008) MEG results. For aspectual coercion versus control stimuli, they observed a right-lateralised effect "which overlapped with previous MEG localisations

[11] See above for a reanalysis-based discussion of the positivity.

of the N400" (Brennan & Pylkkänen, 2008, p. 140). This effect was followed by anterior midline field (AMF) activation similar to that observed by Pylkkänen and McElree (2007) for complement coercion. Brennan and Pylkkänen (2008) assume a pragmatic explanation, in which the earlier activation reflects the computation of an anomalous meaning, followed by a subsequent "meaning shift" (p. 140). In order to assess this potential account further in relation to the present findings, it will be important to examine the electrophysiological correlates of aspectual coercion. A second, but somewhat related, account of our results might assume that there is a division of labour between the N400 and the late positivity in enriched composition, with the N400 reflecting type shifting and related processes (i.e. those processes that are necessitated by a semantic mismatch) and the late positivity reflecting additional discourse-pragmatic updating. According to Schumacher (2011), late positive effects correlating with enriched composition can be functionally interpreted as follows: "the discourse representation is updated in order to reach coherence and to determine the intended meaning on the basis of the wider context" (Schumacher, 2011, p. 214). Possibly, then, enriched composition should not be viewed as a single process but rather as a combination of semantic and pragmatic operations that serve to render an utterance interpretable in a given sentence and discourse context. In addition, our results provide initial evidence that these operations might be subject to interindividual variability.

In summary, our findings provide compelling evidence against an indeterminacy-based account of flexible auxiliary selection. Rather, they suggest that the gradient behaviour of these verbs (e.g. in terms of acceptability judgements) results from a complex interaction of a range of more fine-grained processes during online sentence interpretation. We have suggested that our findings provide initial evidence for three separate processes: (1) a compositional match/mismatch between the properties of the auxiliary and that of the verb, possibly including semantic coercion processes that serve to rectify the mismatch in lexical aspect (reflected in the N400)[12]; (2) further coercion processes, perhaps based on discourse-pragmatics, that vary between individuals and that correlate with an additional increase in the likelihood of a successful interpretation (reflected in the late positivity); (3) a well-formedness categorisation that applies to every utterance encountered in the context of an acceptability judgement task (also reflected in a late positivity, though this effect appears to be qualitatively different from the enrichment positivity described previously).

SUMMARY AND CONCLUSIONS

The present study demonstrated that indeterminacy at the syntax–semantics interface is, in part, processed in a qualitatively similar manner to consistent (rigid) form-to-meaning mappings, while also providing evidence for some degrees of quantitative and qualitative variation. We consistently observed larger N400 effects for mismatches between the aspectual preferences induced by an auxiliary and the aspectual properties of the verb actually encountered. This was even the case for change of state verbs with a low degree of lexical specification for telicity, which showed a correlation between individual item- and participant-based acceptabilities for the auxiliary HAVE and N400 amplitude (lower acceptabilities were associated with more negative-going N400 amplitudes).

[12] Note, however, that the assumption of semantic coercion processes reflected in the N400 does not readily fit with the observation of a negative correlation between N400 amplitude and acceptability (see Footnote 10, for similar arguments against a polysemy-based explanation).

Finally, qualitatively different effects for flexibility versus rigidity were observed in the late positivity time window. Whereas violations of rigid form-to-meaning mappings engendered a late positivity for unacceptable auxiliary selections and, over both auxiliary types, this was also the pattern observed for flexible verbs on a single-trial basis (negative correlation between the interaction between by-item and by-participant acceptability and positivity amplitude), there was an additional positive correlation between late positivity amplitude and by-participant acceptability for HAVE. We suggest that this distinction can be explained in terms of a well-formedness categorisation, which takes place for all utterances in the context of an acceptability judgement task, in comparison to processes of (discourse-based) pragmatic enrichment, which can be applied in the case of more flexible mappings and which appear to vary across individual speakers. In sum, our results indicate that the gradient behaviour of verbs with flexible auxiliary selection is not due to syntactic and/or semantic indeterminacy (underspecification) but rather to additional processes of enriched composition that can apply in the processing of these verbs. We have further suggested that these enrichment processes may also be multidimensional in nature, depending both on subtle differences in the inherent lexical specification of individual verbs and on the propensity of individual speakers for pragmatic enrichment and that these two dimensions are reflected in qualitatively different neural correlates (N400 vs. late positivity).

Manuscript received 30 August 2010
Revised manuscript received 29 June 2012
First published online 5 October 2012

REFERENCES

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baggio, G., Choma, T., van Lambalgen, M., & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience*, *22*, 2132–2140.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, *72*, 32–68.

Bard, E. G., Frenck-Mestre, C., & Sorace, A. (2010). Processing auxiliary selection with Italian intransitive verbs. *Linguistics*, *48*, 325–361.

Bates, D. M. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project*, *5*, 27–30.

Borer, H. (2005). *In name only. Structuring sense, Volume I*. Oxford: Oxford University Press.

Bornkessel, I. (2002). *The Argument Dependency Model: A neurocognitive approach to incremental interpretation (28)*. Leipzig: MPI Series in Cognitive Neuroscience.

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.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., et al. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117*, 133–152.

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

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008b). Unmarked transitivity: A processing constraint on linking. In R. D. Van Valin, Jr. (Ed.), *Investigations of the syntax-semantics-pragmatics interface* (pp. 413–434). Amsterdam: John Benjamins.

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

Bornkessel-Schlesewsky, I., & Schlesewsky, 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.

Brennan, J., & Pylkkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of Aspectual Coercion. *Brain and Language*, *106*, 132–143.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*, 159–168.

Burzio, L. (1986). *Italian syntax: A government-binding approach*. Foris: Dordrecht.

Cennamo, M., & Sorace, A. (2007). Auxiliary selection and split intransitivity in Paduan: Variation, and lexical-aspectual constraints. In R. Aranovich (Ed.), *Split auxiliary systems. A cross-linguistic perspective* (pp. 65–99). Amsterdam: John Benjamins.

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.

de Almeida, R. G. (2004). The effect of context on the processing of type-shifting verbs. *Brain and Language*, *90*, 249–261.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.

Dölling, J. (1995). Ontological domains, semantic sorts and systematic ambiguity. *International Journal of Human–Computer Studies*, *43*, 785–807.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*, 547–619.

Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400–P600 distinction. *Neuropsychologia*, *49*, 574–579.

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

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J. et al. (2009). *Arm: Data analysis using regression and multilevel/hierarchical models*. R package version 1.2-9.01. Retrieved from http://CRAN.R-project.org/package = arm.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–80). Erlbaum: Hillsdale.

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.

Husband, E. M., Kelly, L. A., & Zhu, D. C. (2011). Using complement coercion to understand the neural basis of semantic composition: Evidence from an fMRI study. *Journal of Cognitive Neuroscience*, *23*, 3254–3266.

Keller, F., & Sorace, A. (2003). Gradient auxiliary selection and impersonal passivization in German: An experimental investigation. *Journal of Linguistics*, *39*, 57–108.

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.

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, Germany.

Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*, *22*, 2685–2701.

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*(12), 920–933.

Legendre, G. (2007). On the typology of auxiliary selection. *Lingua*, *117*, 1522–1540.

Legendre, G., & Sorace, A. (2003). Auxiliaires et intransitivite' en francais et dans les langues romanes [Auxiliaries and intransitivity in French and the Romance languages]. In D. Godard (Ed.), *Les langues romanes; problemes de la phrase simple* (pp. 185–234). Paris: Editions du CNRS.

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-semantics interface*. Cambridge, MA: MIT Press.

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.

McElree, B., Traxler, M., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched semantic composition. *Cognition*, *78*, B15–B25.

McFadden, T. (2007). Auxiliary selection. *Language and Linguistics Compass*, *1*, 674–708.

Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research*, *21*, 199–249.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*, 324–346.

Nunberg, G. (1979). The nonuniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, *3*, 143–184.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

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

Otten, M., Nieuwland, M. S., & van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*, 89.

Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*, 157–189.

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. (2006). Underspecification and aspectual coercion. *Discourse Processes*, *42*, 131–155.

Pickering, M. J., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de Almeida. *Brain and Language*, *93*, 1–9.

Piñango, M. M., Winnick, A., Ullah, R., & Zurif, E. (2006). Time-course of semantic composition: The case of aspectual coercion. *Journal of Psycholinguistic Research*, *35*, 233–244.

Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of sentence meaning. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 539–580). London: Elsevier.

Pylkkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience*, *19*, 1905–1921.

R Development Core Team. (2007). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ramchand, G. C. (2008). *Verb meaning and the lexicon. A first phase syntax*. Cambridge: Cambridge University Press.

Schumacher, P. B. (2011). The hepatitis called: Electrophysiological evidence for enriched composition. In J. Meibauer & M. Steinbach (Eds.), *Experimental pragmatics/semantics* (pp. 199–219). Amsterdam: John Benjamins.

Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in native and non-native grammars of Italian. *Second Language Research*, *9*, 22–47.

Sorace, A. (1995). Acquiring argument structures in a second language. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage* (pp. 153–175). Amsterdam: John Benjamins.

Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, *76*, 859–890.

Sorace, A. (2004). Gradience at the lexicon-syntax interface: Evidence from auxiliary selection. In A. Alexiadou, M. Everaert, & E. Anagnostopoulou (Eds.), *The unaccusativity puzzle* (pp. 243–268). Oxford: Oxford University Press.

Sorace, A. (to appear). *Gradience in split intransitivity*. Oxford: Oxford University Press.

St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, *9*, 776–787.

Traxler, M., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, *47*, 530–547.

Van Berkum, J. J. A., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 443–467.

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*, 1211–1224.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*, 67–82.

Van Valin, R. D., Jr. (1990). Semantic parameters of split intransitivity. *Language*, *66*, 221–260.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantc integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1271–1288.

APPENDIX 1

Alternative mixed-effects model fits using centred acceptability ratings.

Best-fitting models for the N400 and late positivity time windows using centred acceptability ratings are shown in Tables A1 and A2, respectively.

TABLE A1

Parameter values for fixed effects in the best-fitting linear mixed-effects models of mean amplitude values for the CH-STATE-UN verbs in the N400 time window (380–530 ms) using *centred acceptability ratings* and associated confidence intervals and *p*-values. Confidence intervals and *p*-values were obtained by MCMC sampling (Baayen et al., 2008). Note that only data from the posterior electrodes entered this model (see the main text) and that, for the factor auxiliary type, BE was the reference level. For reasons of readability, only significant effects are reported (note that the separate analysis for BE revealed no significant effects)

	Estimate	*95% CI-lower*	*95% CI-upper*	*p-Value*
Global analysis				
Intercept	1.37	0.86	1.85	<.0001
AUX-(have)	0.31	0.14	0.50	<.001
ITM-ACC	1.30	0.76	1.87	<.0001
AUX-(have):ITM-ACC	−1.10	−1.95	−0.27	<.02
AUX-(have):ITM-ACC:SUBJ-ACC	5.72	3.31	8.35	<.001
HAVE				
SUBJ-ACC	−2.66	−5.18	0.17	<.06
SUBJ-ACC:ITM-ACC	6.75	5.15	8.50	<.0001

SUBJ-ACC, participant-based acceptability; ITM-ACC, item-based acceptability; AUX, auxiliary type.

TABLE A2

Parameter values for fixed effects in the best-fitting linear mixed-effects models of mean amplitude values for the CH-STATE-UN verbs in the late positivity time window (750–900 ms) using *centred acceptability ratings* and associated confidence intervals and *p*-values. Confidence intervals and *p*-values were obtained by MCMC sampling (Baayen et al., 2008). Note that only data from the posterior electrodes entered this model (see the main text) and that, for the factor auxiliary type, BE was the reference level. For reasons of readability, only (at least marginally) significant effects are reported (note that the separate analysis for BE revealed no significant effects)

	Estimate	*95% CI-lower*	*95% CI-upper*	*p-Value*
Global analysis				
AUX-(have):SUBJ-ACC	1.53	0.56	2.47	<.001
SUBJ-ACC:ITM-ACC	−3.28	−4.52	−2.01	<.0001
HAVE				
SUBJ-ACC	3.06	0.61	5.75	<.02

ACCEPT-SUBJ, participant-based acceptability; ACCEPT-ITM, item-based acceptability; AUX, auxiliary type.

N400 time window. In order to examine the underlying cause of the interaction of by-participant and by-item acceptability for HAVE, which suggests that the effect of by-item acceptability on ERP amplitude is dependent on by-participant acceptability, we performed a median split on by-participant acceptability and then fit separate models including by-item acceptability for the two groups of participants (i.e. the "high-acceptability group", those judging CH-STATE-UN with HAVE with above-median acceptability, and the "low-acceptability group", those judging CH-STATE-UN with HAVE with below-median acceptability).

These revealed an effect of by-item acceptability only for the high-acceptability group (model estimates: intercept 1.34; 95% CI: 0.58–2.14, $p < .01$; by-item acceptability 1.95; 95% CI: 0.73–3.11; $p < .01$), but not for the low-acceptability group (model estimates: intercept 1.82; 95% CI: 1.06–2.67, $p < .001$; by-item acceptability 0.19; 95% CI: -1.13 to 1.46; $p > .78$). Thus, the reduction of N400 amplitude with increasing by-item acceptability is driven primarily by those participants who have a higher tendency to accept CH-STATE-UN verbs with HAVE.

Time window 2. As for time window 1, we assessed the cause of the by-participant and by-item acceptability interaction in the global analysis for time window 2 by performing a median split on by-participant acceptability (in this case, over both auxiliaries, since there was no three-way interaction with AUX) and fitting separate models including by-item acceptability for the high- and low-acceptability groups. Only by-item acceptability was included in the models (i.e. not AUX or the interaction between the two factors) because of the absence of the three-way interaction in the global analysis. These separate analyses revealed effects of by-item acceptability for both groups, though the effect was stronger in the high-acceptability group (model estimates: intercept 0.51; 95% CI: -0.11 to 1.18, $p > .1$; by-item acceptability -1.91; 95% CI: -2.27 to -1.52; $p < .0001$) than in the low-acceptability group (model estimates: intercept 0.12; 95% CI: -0.36 to 0.62, $p > .6$; by-item acceptability -0.50; 95% CI: -0.85 to -0.14; $p < .01$). In both groups, higher acceptability ratings correlated with smaller positivity amplitudes.