When verbs have bugs: Lexical and syntactic processing costs of split particle verbs in sentence comprehension

Anna Czypionka[a,b] and Felix Golcher[c], Joanna Błaszczak[b], Carsten Eulitz[a]

[a]Department of Linguistics, University of Konstanz [b]Institute of English Studies, University of Wrocław , [c]Department of Linguistics, Humboldt-Universität zu Berlin

ARTICLE HISTORY
Compiled October 15, 2018

ABSTRACT
German particle verbs consist of a base and a particle, two constituents which occupy separate positions in main clauses, but share one lexical entry. It is still unclear if the combination of particles and bases during sentence comprehension is lexical, syntactic or dual in nature. Using behavioural and ERP measurements, we investigated lexical access and sentence integration of split particle verbs in German two-argument sentences. Our results show that the integration of split particle verbs violating sentence structure or lexical constraints leads to both lexical and syntactic processing difficulty. This extends earlier comparable findings reporting only lexical access difficulties, and suggests that the parse is not immediately abandoned upon encountering a nonexistent particle verb. The integration of grammatical particle verbs assigning lexical case did not lead to measurable processing difficulties. We discuss the impact of this finding for current accounts of the role of lexical case marking in sentence comprehension.

KEYWORDS
sentence comprehension; particle verbs; lexical access; transitivity; case marking; ERP; German

1. Introduction

During sentence comprehension, most words we encounter correspond to one lexical unit - and in general, one lexical unit is represented by one word. One interesting exception are discontinuous words, multiple words that share one lexical entry. An example for these are particle verbs, verbal compounds like *look up*, which are especially common in Dutch and German.

Particle verbs like German *anhören* ('to listen to') consist of a base verb (*hören*, 'to hear') and a separable prefix, the particle (*an*). The verb base usually is a verb that also bears an independent meaning, while particles are often homologuous to prepositions or adverbs. In German subordinate clauses, particle verbs occur as one verbal unit at the

CONTACT Anna Czypionka. Email: anna.czypionka@uni-konstanz.de

sentence-final position, in keeping with the underlying SOV word order of German: ... *dass Peter den riesigen Kuchen aufisst* ('... that Peter finishes the giant cake', literally: that Peter the giant cake up-eats). In German main clauses with simple tense (present or preterite), the finite verb occurs in the second position, revealing number, person and tense information. In the case of particle verbs, however, only the base verb is moved to the second-constituent position. The particle is split from the base and remains in the sentence-final position: *Peter isst den riesigen Kuchen auf* ('Peter finishes the giant cake', literally: Peter eats the giant cake up.). It is thus possible that the verb in a German main clause will turn out to be a particle verb once the last word in the sentence has been recognised. This possibility is by no means certain, since many base verbs also occur as standalone nonseparable simple verbs without a particle. This means that in German main clauses with simple tenses, the full lexical information about the verb is only accessible at the clause-final position, and that its semantic and syntactic properties are only revealed after all potential arguments have been accessed.

How does the parser deal with these split words that surface in two different positions in the sentence? How does it deal with the potential, but uncertain dependency that is only resolved at the clause-final position? Which processes are triggered when a verb's predicted semantic and syntactic properties are confirmed - or changed - upon encountering the particle?

Research on particle verbs in sentence comprehension has focused on the semantic ambiguity introduced by potential base verbs, on the expectations built up by the parser upon encountering potential base verbs, on the plausibility of certain base-particle combinations, and on the lexical status of particle verbs (Cappelle, Shtyrov, & Pulvermüller, 2010; Isel, Alter, & Friederici, 2005; Piai, Meyer, Schreuder, & Bastiaansen, 2013; Urban, 2001, 2002). So far, little attention has been paid to the *syntactic* ambiguity in base-particle sentences and no correlates of syntactic processing difficulty have been reported for sentences with nonexistent base-particle verbs. In this paper, we set out to investigate the comprehension processes in German transitive clauses at the clause-final particle, when the combination of particles and base verbs are revealed, and when lexical access to the verb is finally possible. We will compare the processing of 'easy', standard nominative-accusative-assigning base-particle combinations with different deviations from this 'standard' pattern:

- Illegal base-particle combinations forming non-existing particle verbs. The processing of these verbs should reflect difficulties in lexical access and subsequent breakdown of the parse.
- Legal base-particle combinations forming intransitive, one-place particle verbs. While there should be a lexical entry for these base-particle-combinations, they should lead to syntactic processing difficulties if the particle leading to an intransitive verb is encountered at the end of a clause containing two potential argument NPs.
- Legal base-particle combinations forming transitive two-place particle verbs that assign nominative-dative instead of nominative-accusative. While both case marking patterns occur in German two-place verbs, lexical nominative-dative case marking is more rare, and has been suggested to lead to a mild enhancement of lexical processing load (Bader, Meng, & Bayer, 2000; Bayer, Bader, & Meng, 2001).

These comparisons will allow us to investigate different types of processing difficulties that can arise upon encountering a verb-final particle.

In the following, we will give a short background on the processing of particle verbs. We will conclude the Introduction by taking up our research question again, in more detail, explaining how we extend earlier research, and by formulating predictions for the different stimulus conditions.

1.1. Background

German particle verbs are compound verbs, consisting of a particle and a base. In main clauses with simple tense, these compounds are realised as multiple words (McIntyre, 2007), with the base verb occurring in V2, and the particle occurring in the clause-final position, thereby forming a syntactic and lexical dependency. However, this dependency is uncertain, since many base verbs can also be realised without a particle in similar contexts. One example is the verb *hören* (to hear), which can stand on its own, but can also serve as a base verb in base-particle combinations. Example 1 is an illustration of a main clause with a simple verb, and with the same simple verb serving as the base for a particle verb. [1]

Example 1 Examples of base verb with and without particle.

(A) *hören*, no particle:
Peter **hört** Musik auf seinem alten Kassettenrekorder.
Peter hear music on his old tape.recorder
'Peter is listening to music on his old tape recorder.'

(B) *anhören* (to listen to):
Peter **hört** eine Kassette auf seinem alten Kassettenrekorder **an**.
Peter hear a tape on his old tape.recorder PART
'Peter is listening to a tape on his old tape recorder.'

Particle verb formation is highly productive, and it is possible to combine base verbs with particles to form new, perhaps surprising but immediately understandable particle verbs. A corpus search can easily yield over thirty potential particles for one base verb. Assessing the frequency of specific base-particle combinations using currently available corpora is difficult. This is because particle verbs are only recognised in subordinate clauses or in main clauses with an auxiliary verb, when they occur as one orthographic unit in the clause-final position, and therefore, the absolute lemma frequencies of particle verbs are systematically underestimated (while the absolute lemma frequencies of potential base verbs are overestimated, see also Smolka, Preller, and Eulitz 2014, p. 33, for a short discussion of this issue).

Although the constituent parts of particle verbs can surface at different sentence positions (McIntyre, 2007), a particle verb is usually assumed to have one lexical entry (Cappelle et al., 2010; Jackendoff, 2002). Just like other types of compounds, particle verbs can be semantically transparent, with a close semantic relatedness between the meaning of the base verb and the particle verb (as in the case of *hören* and ***an****hören*), or semantically opaque (like ***auf****hören*, 'to stop'). Both semantically transparent and opaque particle verbs have been argued to be lexically associated with their base verbs, regardless of the semantic relationship between the base verb and the particle-base combination (see Smolka, Komlósi, and Rösler 2009; Smolka et al. 2014 for German prefix and particle verbs, and Zwitserlood, Bolwiender, and Drews 2005 for Dutch particle verbs[2]). German complex verbs (including both nonseparable prefix verbs and separable particle verbs) are lexically associated with their base verbs, regardless of semantic relatedness (see Smolka et al. 2009, 2014). This indicates that the lexical entries of complex verbs are represented via their bases and the particles or prefixes, irrespective of whether the complex verb is semantically transparent or opaque. While particle verbs and prefix verbs have been used in a number of studies investigating single word and compound processing (e.g., Smolka et al., 2009, 2014; Zwitserlood et al., 2005), and particle verbs are often used along with simple verbs in sentence comprehension experiments, there are only a handful of studies explicitly investigating the processing of particle verbs in sentence comprehension. We will give an overview over these studies here to illustrate a number of open issues.

Comparisons of existing and non-existing base-particle combinations In a series of EEG studies, Urban (2001, 2002) monitored the processing of German main clauses containing split verbs. Measurements at the clause-final particle revealed an enhanced N400 for illegal base-particle combinations (when the verb in V2 was a simple intransitive verb that does not combine with a particle), compared to legal base-particle combinations (i.e., verbs in V2 that occur both as stand-alone simple intransitive verbs and bases for particle verbs). For legal compared to illegal conditions, an enhanced P300 was found. This latter finding was dependent on the behavioural task employed, and occurred with grammaticality judgments, but not with a probe detection task.[3]

Isel et al. (2005) monitored the ERPs to clause-final particles in auditorily presented German sentences, comparing between legal conditions (e.g., *(an)lächeln*, 'to smile (at)') and illegal conditions (i.e., with verbs in V2 that never combine with particles, e.g. *(*an)nennen*, 'to name (*at)'. They found late frontal negativities for illegal compared to legal conditions, irrespective of sentence prosody. The authors interpret this late frontal negativity as reflecting a "checking procedure for attempting to combine together the particle and the verb" (Isel et al., 2005, p.161). Only when sentence prosody marked the presence of a sentence-final particle, an additional N400 was found on the particle for illegal verb-particle combinations. No P600 was found in sentences with nonexistent verb base-particle combinations, irrespective of sentence prosody. The results of Isel et al. (2005) suggest that the expectations of finding a particle (as built up by prosody) play an important part in the comprehension of base-particle combinations; and that when a particle is expected, an illegal combination will lead to an enhanced N400.

In an MEG study using English verb bases followed by particles forming existing or non-existing verb-particle combinations, Cappelle et al. (2010) found evidence that particle verbs are accessed as a single lexical unit, and not combined via rule-based processes. Importantly, this finding holds for both semantically opaque and transparent particle verbs. The authors interpret their findings as showing that particle verbs have a single lexical entry, and that their meaning is not composed anew upon every new encounter. The authors also conclude that this supports the assumption that the mental lexicon in fact stores more than just the idiosyncratic parts of grammatical knowledge. Extending these earlier findings, Hanna, Cappelle, and Pulvermüller (2017) report EEG data from the processing of split particle verbs in a short sentence context. The authors conclude that even in these contexts, split particle verbs are accessed as one lexical item, and do not need to be syntactically combined. This finding holds for both semantically opaque and transparent particle verbs.

Piai et al. (2013) used visually presented Dutch sentences with split particle verbs in two EEG studies. In their first experiment, they found enhanced left-anterior negativities on verbs that serve as potential bases for particle verbs when compared to verbs that never combine with particles. The amplitude of these negativities did not depend on the number of particles that combine with a potential base; the authors conclude that this indicates an underlying syntactic rather than lexical process, and propose an increased working memory load for potential bases. In their second experiment, Piai et al. (2013) manipulated the semantic plausibility of the sentences by combining bases with particles that led to semantically plausible, semantically implausible, or nonexistent base-particle combinations. For this experiment, the authors found a graded N400 on the sentence-final particle, with smaller negativities for existing and plausible particle verbs, larger negativities for existing, but implausible particle verbs, and most negative values for nonexistent verb-particle combinations. Interestingly, no P600 was found for nonexistent verb-particle combinations. This leads to the question if the parser immediately abandons the buildup of syntactic structure after having encountered a verb without a lexical entry, and if the failure is purely lexical (i.e., not finding the lexical entry), but not syntactic (i.e., failing to establish a dependency between the base and a potential particle). From the results of both experiments, the authors conclude that the lexical entry of split particle verbs is accessed twice during sentence comprehension: once upon encountering the base, the other time upon encountering the particle.

Taken together, the studies cited above have found an N400 on clause-final particles for nonexistent compared to existing base-particle combinations. This N400 was more pronounced than an N400 engendered by an existing, but semantically implausible or unexpected base-particle combination (Piai et al., 2013). This leads to the question if the N400 found on illegal or unexpected particles represent difficulties in lexical access, semantic processing, or a mix of various factors. Interestingly, none of the studies has reported a P600 for nonexistent base-particle combinations. This, however, could arguably be expected: Once a sentence turns out to not contain an existing verb, it should lead to enhanced syntactic processing load (due to difficulties in representation building, or integration of words into the sentence context), which is often reflected in an enhanced P600 (Friederici, Pfeifer, & Hahne, 1993; Hagoort, 2003; Hagoort, Brown, & Groothusen, 1993; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005). In addition, the P600 has also been argued to reflect enhanced semantic processing load, especially in sentences with difficulties in argument role assignment or poor lexico-semantic fit between potential arguments and verbs (Hoeks, Stowe, and Doedens 2004; Kim and Osterhout 2005; Kuperberg, Kreher, Sitnikova, Caplan, and Holcomb 2007; see also Brouwer, Fitz, and Hoeks 2012 for an account of ERP components reflecting lexical access and integration into the sentence context).

Comparison of existing particle verbs assigning different cases As outlined above, when the sentence-final particle is revealed and combined with the base verb, a sentence can become semantically implausible, syntactically anomalous, or both. Furthermore, a nonexistent base-particle combination can cause increased lexical processing difficulty as the parser searches for a nonexistent lexical entry. This lexical processing difficulty has been shown to elicit an increased N400. In German there is an additional and more subtle morphosyntactic and semantic manipulation that can be introduced via the sentence-final particle: The particle can change the object case of two-place verbs from the standard accusative to the nonstandard lexical dative. This manipulation could be assumed to enhance lexical processing cost without introducing semantic or syntactic violations. Here, we will give a short outline of the characteristics of this manipulation, before explaining how we will use it for our research.

The German verb *hören* ('to hear') may combine with particles resulting in verbs assigning standard nominative-accusative case (***anhören***, *Ich höre die Geschichte* ***an***., 'I listen to the.ACC story(.ACC), or non-standard nominative-dative case (***zuhören***, *Ich höre der Geschichte* ***zu***., 'I listen to the.DAT story(.DAT)). Due to widespread case syncretism on many German nouns, some accusative and dative arguments look identical, so that grammatical sentences are possible where the case marking is only revealed on the sentence-final verb or particle (*Ich höre Sängerinnen an / zu*, 'I listen to singer.FEM.PL.(.ACC/DAT)'). Unlike structural accusative, dative is considered an idiosyncratic or lexical case for objects of two-place verbs (Czepluch, 1996; Haider, 1993, 2010; Woolford, 2006), and signals non-prototypically transitive argument semantics (Blume, 2000; Dowty, 1991; Grimm, 2010). Revealing that a verb assigns lexical dative instead of structural accusative should have consequences for processing, since there are a number of processing differences between nominative-accusative and nominative-dative verbs. Compared to sentence-final NOM-ACC verbs, NOM-DAT verbs cause increased processing costs when the case marking on the arguments is morphologically ambiguous. These increased processing costs manifest as increased reaction times in behavioural experiments (Bader et al., 2000), and as N400 effects at the clause-final verb in EEG studies (Hopf, Bader, Meng, & Bayer, 2003; Hopf, Bayer, Bader, & Meng, 1998). Unlike NOM-ACC verbs, NOM-DAT verbs come with unmarked subject-object and object-subject verb orders (Bader, 1996; Bornkessel, McElree, Schlesewsky, & Friederici, 2004), and show an attenuated influence of argument animacy contrasts (Czypionka, 2014; Czypionka, Spalek, Wartenburger, & Krifka, 2017). To account for the increased processing load for nominative-dative verbs, it has been suggested that the processing of lexical case marking is in fact a two-step process (Bader et al., 2000; Bayer et al., 2001). The first step is a mild reanalysis of syntactic structure. The second step is reaccess of the lexical entry of the *object* to check if lexical dative case is morphologically licensed (which is unneccessary with structural accusative case); it is this step that is thought to be reflected in the N400 found in ERP studies on the processing of NOM-DAT verbs. This proposal is also known as the 'Lexical Reaccess Hypothesis'. Another proposal holds that processing differences between accusative- and dative-assigning two-place verbs reflect semantic-thematic differences between both verb types (Bornkessel-Schlesewsky & Schlesewsky, 2006, 2009, 2013). However, the studies cited above used a mix of simple and particle verbs in their stimulus material. This is problematic because simple and particle verbs differ in their morphological and syntactic complexity, and because there is currently much debate about the argument status of the objects of particle verbs (see McIntyre 2015 for a comprehensive overview of the different positions). This makes it difficult to interpret the results and to attribute them to syntactic, semantic or lexical processing differences between dative and accusative. A recent study suggests that at least some of the processing differences between nominative-accusative and nominative-dative verbs hold only for simple verbs, but not for particle verbs, when these particle verbs are presented as one orthographical unit in the clause-final position (Czypionka & Eulitz, 2016, to appear). This finding is especially problematic for the Lexical Reaccess Hypothesis of dative processing, since dative for direct objects is a lexical case in both simple and particle verbs, and should elicit lexical reaccess to case-ambiguous objects upon encountering the clause-final verb (reflected in an enhanced N400). Currently, there is no study comparing directly the processing of *split* particle verbs assigning nominative-accusative and nominative-dative. A study like this could reveal if a particle-induced change from accusative to dative objects is reflected in an enhanced N400; this finding would strengthen the idea that the processing of lexical dative is indeed associated with higher lexical processing load.

To sum up the preceding overview of the literature, previous studies have shown that verbs in V2 serving as potential bases lead the parser to expect a sentence final particle (Isel et al., 2005; Piai et al., 2013). A clause-final particle leading to a nonexistent or semantically implausible base-particle combination is reflected in an enhanced N400, usually taken to reflect increased processing costs for lexical access (and possibly semantic processing difficulty) (see Isel et al. 2005; Piai et al. 2013; Urban 2001, 2002). So far, no indications of a P600 have been found for nonexistent or semantically implausible base-particle combinations, although it could be argued that a nonexistent verb should lead to syntactic as well as lexical and semantic integration difficulties. Thus, the question remains whether nonexistent or implausible base-particle combinations in sentences represent a purely lexical problem for the parser, or if they can be argued to also increase syntactic processing load.
Apart from the contrast between grammatical and ungrammatical/implausible base-particle combinations, there is another contrast that could be expected to increase lexical processing load. This is the contrast between either dative or accusative objects of two-place verbs (Bayer et al., 2001; Hopf et al., 2003, 1998). Importantly, this contrast is not associated with syntactic or semantic violations. However, recent findings have led to doubt if there are in fact differences associated with the processing of nominative-dative and nominative-accusative assigning particle verbs (Czypionka & Eulitz, to appear). This is problematic for current explanations of processing differences between accusative and dative, because dative assigned by particle verbs should also be a lexical case associated with enhanced lexical processing load. It remains open whether the processing of nominative-dative will lead to enhanced lexical processing load if this non-standard case marking pattern is revealed at the clause-final particle, and if this enhanced lexical processing load will be comparable with problems in lexical access caused by outright morpholexical violations (caused by nonexistent base-particle combinations).

Research Goals and Hypotheses Previous studies investigating the processing of split particle verbs have mainly dealt with semantic and lexical, rather than syntactic changes that can be introduced by the clause-final particle. In the current study, we set out to shed further light on the processing of split particle verbs, and to monitor the influence of *syntactic* changes that may be caused by combining a base verb with different particles.[4] To this end, we investigate the processing of German two-argument sentences containing identical base verbs. These base verbs are combined with four different types of particles, leading to four different types of verbs:

- Existing base-particle combinations assigning nominative-accusative case (i.e., the standard case marking pattern in German.); this condition will serve as the baseline that the other conditions are compared to.

- Non-existing or illegal base-particle combinations; these verbs do not have a lexical entry, therefore, sentences in this condition are revealed to not contain an existing verb at the clause-final particle.
- Existing base-particle combinations leading to an intransitive verb. This combination has not been used in previous experiments on particle verb processing. These verbs have a lexical entry, but the combination of base and particle is unexpected given the syntactic and semantic sentence context. Upon encountering the clause-final particle, sentences in this condition are revealed to contain one NP that cannot be assigned an argument role.
- Existing base-particle combinations leading to a transitive verb that assigns nominative-dative (i.e., the lexical case marking pattern in German). These verbs have a lexical entry, and sentences in this condition are grammatically and semantically well-formed. It is yet unclear if they should be associated with enhanced lexical processing load due to lexical reaccess to the object (following Bayer et al. 2001).

We will monitor the processing of these four conditions in three different experiments using different experimental paradigms. Acceptability ratings using Magnitude Estimation tasks will reveal systematic acceptability differences between conditions. In a self-paced reading time study, we will check for general differences in processing load between conditions, visible at the clause-final particle. Finally, an EEG study will allow us to check the specific predictions formulated for EEG measurements at the clause-final particle, and to discuss our findings with respect to the literature. For data analysis, we will perform a row of planned comparisons:

(1) accusative-illegal: By comparing the baseline condition to the one containing nonexistent words, we will measure difficulties caused by lexicon search for a nonexistent lexical entry. We will also monitor processing difficulties caused by sentences that turn out to not contain an existing verb.

(2) accusative-intransitive: By comparing the baseline condition to one containing an existing, but syntactically (and also semantically) non-matching word, we will measure difficulties caused by unexpected particles, and possibly syntactic and semantic integration difficulties.

(3) intransitive-illegal: By comparing between both types of ungrammatical conditions, we will assess if there are qualitative or quantitative differences between both types of violation (nonexistent words compared to existing words that do not match the sentence context).

(4) accusative-dative: By comparing between both grammatical conditions, we will test the prediction of the Lexical Reaccess Hypothesis that dative should lead to mildly enhanced lexical processing load.

We will describe the construction of the language material in more detail in the following section, before we report the three experiments.

2. Language Material

As mentioned in the Introduction, there is no finite number of German particle verbs, since particle verb formation is highly productive. For language material construction, we collected 56 verbs from the DWDS corpus (www.dwds.de) that can serve as potential bases for different types of particle verbs - intransitive, transitive and often also ditransitive verbs. When occurring without a particle, these verbs were intransitive or transitive verbs. Of those 56 verbs, 25 combine with different particles forming both accusative- and dative-assigning two-place verbs. [5] From the 25 base verbs allowing the formation of accusative- and dative-assigning particle verbs and intransitive verbs, 17 were chosen as base verbs in our study. Base verbs were chosen so that they combined with particles to form existing accusative-assigning, dative-assigning or intransitive verbs, and that both types of two-place verbs were semantically appropriate in comparable sentence contexts.

Each base verb was combined with four different particles, leading to four different verb types and conditions:

accusative: existing particle verb, two-place, NOM-ACC case marking. Sentences with this particle verb are grammatical and serve as the baseline condition.

illegal: nonexistent particle verb, no-place, no case marking.

intransitive: existing particle verb, one-place, NOM case marking.

dative: existing particle verb, two-place, NOM-DAT case marking.

The nonexistence of the particle verbs in the illegal condition was assessed with a search in the dlexDB data base (Heister et al., 2011). No hits were found for any of the particle verbs used in the illegal condition. For two particle verbs, the dlexDB corpus gave spurious entries associated with frequencies[6]. For five existing intransitive particle verbs (that were used in nine sentence quartets), we did not find associated frequencies in the dlexDB database: ***mit**spionieren* 'to spy along'; ***weiter**handeln* 'to continue trading/haggling/dealing; to continue acting (in a certain way)'; ***weiter**stellen* 'to continue putting' [implicit: stuff in places]; ***mit**stehen* 'to stand along' (e.g., together with a friend at a bus stop, smokers in front of a restaurant); ***los**jagen* 'to run off'. We checked their existence with a search in the DWDS corpus, finding hits for all of them in the DWDS data base; this suggests that these verbs do indeed exist, but have not been annotated for the corpus yet.

The resulting verbs of all four conditions were used in sentences with the following structure:

subject | base verb | object | modifying PP/NP | adverb | particle | spillover region

For all sentences, subjects were animate and objects were inanimate. Arguments were chosen to fit the semantic selectional restrictions of the NOM-ACC and NOM-DAT verbs in the two grammatical conditions. Subjects and objects were bare plural NPs without morphologically overt case marking. This is possible because many nouns are morphologically ambiguous with respect to case, especially in their plural forms, due to widespread case syncretism in German. The spillover region was a four-word subordinate clause starting with a complementiser (*weil* 'because', *damit* 'for'). An example of a stimulus quartet is given in Example 2.

Example 2 Example of a typical sentence quartet. Note that case morphology is not marked overtly on the arguments. Base verbs and particles are marked in bold letters. The verb type resulting from the specific base-particle combination is given in the condition name, and additionally marked in the glosses of the particle.

(A) accusative:

Polizisten **hören** Unterhaltungen von Kriminellen
policeman.PL.(NOM) hear.3.PL conversation.PL.(ACC) of criminals
heimlich **ab**, damit sie Bescheid wissen.
secretly PART-ACC so.that. they about.it know

'Policemen secretly wiretap criminals' conversations so that they know what is going on.'

(B) illegal:

Polizisten **hören** Unterhaltungen von Kriminellen heimlich
policemanPL.(NOM) hear.3.PL conversation.PL of criminals secretly
über, damit sie Bescheid wissen.
PART-ILL so.that. they about.it know

'Policemen secretly VERB? criminals' conversations so that they know what is going on.'

(C) intransitive:

Polizisten **hören** Unterhaltungen von Kriminellen heimlich
policemanPL.(NOM) hear.3.PL conversation.PL of criminals secretly
weg, damit sie Bescheid wissen.
PART-INTR so.that. they about.it know

*'Policemen secretly turn a deaf ear criminals' conversations so that they know what is going on.'

(D) dative:

Polizisten **hören** Unterhaltungen von Kriminellen
policemanPL.(NOM) hear.3.PL conversation.PL.(DAT) of criminals
heimlich **zu**, damit sie Bescheid wissen.
secretly PART-DAT so.that. they about.it know

'Policemen secretly listen to criminals' conversations so that they know what is going on. '

Following this pattern, we constructed 40 sentence quartets. For the 40 sentence quartets, 25 accusative verbs, 21 dative verbs, 25 intransitive verbs and 30 nonexistent particle verbs were used. No individual particle verb was repeated more often than 4 times. The particle verbs used in the stimulus quartets were constructed from 16 different base verbs (*hören, halten, sehen, fahren, spionieren, handeln, stellen, rennen, jubeln, stehen, laufen, springen, schauen, gucken, jagen, geben*) and 41 different particles (*an, zu, auf, über, ab, weg, hin, aus, stand, gegen, neben, mit, rum, nach, ran, wieder, entgegen, halt, um, voran, dar, bereit, klar, runter, zuwider, weiter, herunter,* *wett, sicher, inne, her, hoch, durch, gegenüber, fest, nieder, entlang, fort, los, acht*). For four of the 40 sentence quartets, the sentences in the different conditions were not completely identical in the pre-particle-region; in these quartets, the pre-particle adverbs or direct objects were varied to achieve natural-sounding sentences in the grammatical conditions. A full overview of the stimulus material is given in the Supplementary Material, the four special sentences are explicitly marked. [7] Before the start of the experiments, the stimuli were checked in a number of pretests: a sentence completion task to assess if the unexpected items were truly unexpected, and a frequency check for the existing two-place verbs. [8]

Sentence completion task Before the start of the experiments, we ran a sentence completion task, using 36 of the 40 items (so that all of the 16 different base verbs appeared at least once, and no base was used more often than three times). This was done to assess if the base-particle combinations used in the ungrammatical conditions (intransitive and illegal) would not be supplied by participants. In addition, the sentence completion task allowed us to assess if there were different expectations for particles leading to nominative-accusative or nominative-dative verbs in our stimuli, i.e., whether the accusative or the dative condition were more plausible in our stimulus set. 54 monolingual native speakers of German participated in the sentence completion task. The mean age of participants was 21.5 years ($SD = 3.0$). 13 participants were male. The task was performed as a pen-and-paper test as part of a course assignment. All sentences were printed without the post-particle spill-over region. The space for the particles was left blank, so that participants saw sentences like *Polizisten hören Unterhaltungen von Kriminellen heimlich* ________. The 36 critical sentences were interspersed with 40 filler sentences with blanks in various sentence positions. These filler sentences were not designed to elicit verbal particles. In the sentence completion task, 92.2% of sentences were completed with a particle leading to a transitive verb. We therefore conclude that a particle is the expected continuation in this position. Of the sentences completed with a particle, 58.2% were completed with ACC continuations, and 41.8% with DAT continuations. This allows the conclusion that both dative and accusative particle verbs are expected at the clause-final position in sentences like our stimuli. It also suggests that although both types of continuations are provided, expectations for accusative particle verbs are slightly higher than those for dative particle verbs in these sentences. [9]

Frequencies for existing verbs As outlined in the Introduction, frequencies for particle verbs are systematically underestimated (and those for potential base verbs are overestimated) in German corpora. Nonetheless, we checked the existing verb types for frequency, assuming that this flaw should affect all existing particle verbs to a comparable extent. For the existing verbs (excluding the intransitive verbs that did not have hits in the dlexDB database), there was no effect of condition on lemma frequency ($F(2, 63) = 1.7$, $p > .1$). A planned comparison between accusative- and dative assigning verbs again showed no effect of condition on lemma frequency ($F(1,44)$=1.26, p>.2).

The stimulus material outlined here was used in three different experiments: An acceptabiltiy rating study, a self-paced reading time study, and an EEG study. We will formulate predictions for the experimental outcomes of all conditions in the sections describing the individual experiments.

3. Experiment 1: Magnitude Estimation

In the first experiment, the stimuli were rated for acceptability in a Magnitude Estimation task (Bader, 2012). Apart from checking our own intuitions about the acceptability of our conditions, this was done to assess subtle differences in acceptability between both grammatical conditions (accusative and dative), and between both ungrammatical conditions (intransitive and illegal). We predict the following outcomes for the different comparisons:

- accusative vs. illegal: Illegal sentences should receive lower ratings than accusative sentences, since illegal sentences contain a lexical violation (i.e., a word that does not exist).
- accusative vs. intransitive: Intransitive sentences should receive lower ratings than accusative sentences, since intransitive sentences contain a syntactic (and arguably also a semantic/thematic) violation.
- intransitive vs. illegal: Both conditions should be judged as unacceptable. We do not predict a marked differefor this comparison, nce between both conditions, since we assume that the difference in violations is qualitative rather than quantitative.
- accusative vs. dative: Both conditions should be judged as acceptable. Given the results of the sentence completion task, it is possible that dative conditions would receive slightly lower ratings than accusative conditions.

3.1. Material and Methods

Participants 59 participants were tested. Participants were recruited via the SONA systems participant data base of the University of Konstanz. All participants spoke German as their only native language, and reported no known neurological or reading-related problems. All participants gave written informed consent. Participants received 2 Euros compensation. Three participants were excluded from the analysis, one for a lack of cooperation. The other two participants obviously misunderstood the instructions, consistently rating ungrammatical sentences higher than grammatical ones, and using a much smaller value range for their judgments than the remaining participants. The remaining 56 participants were between 18 and 30 years old (mean age = 22.7 years, SD = 3.0). 11 participants were male.

Language Material The Language Material in this study is outlined in section 2. For each participant, one sentence from each of the 40 critical quartets was presented. The 40 critical quartets were interspersed with 80 filler sentences belonging to eight different conditions, three of which were ungrammatical and five of which were grammatical. Each participant saw 120 sentences.

Procedure The experiment was run in the Psycholinguistics Lab of the University of Konstanz. Stimuli were presented on a 17″ cathode ray tube monitor (Sony Trinitron Multiscan G400), connected to a Fujitsu personal computer. Stimuli were presented and ratings were recorded in Linger (Rohde, 2003). All sentences were rated relative to a reference sentence. The reference sentence was *Die Mitarbeiter haben dass der Chef Probleme hat wohl nicht sofort bemerkt.* ("Apparently, the coworkers did not notice right away that the boss was having problems.'). The acceptability for this reference sentence was set to 50 (following Bader 2012). Participants were instructed to rate sentences with higher acceptability with higher values, and sentences with lower acceptability with lower values. The lower limit for (bad) ratings was 1; there was no upper limit to the possible ratings. Before the start of the actual experiments, participants rated five practice sentences.

Data analysis Data preparation and analysis were performed in R (R Core Team, 2017). In 19 cases, participants gave a rating of 0. These ratings were removed from the data analysis. For further analysis, we z-scaled the ratings per subject. Figure 1 shows a density plot of the rescaled data per condition. The scaled data are not normally distributed, showing heavy tails instead, with the grammatical conditions accusative and dative more on the left and the ungrammatical condtions intransitive and illegal more on the right. Because of this non-normal distribution of the data, it is necessary to use a non-parametric test.

Based on our initial hypotheses, we performed four specific comparisons: (1) accusative vs. illegal, (2) accusative vs. intransitive, (3) intransitive vs. illegal, and (4) accusative vs. dative. For each of these comparisons, we ran Wilcoxon signed rank tests (Wilcoxon, 1945). These tests were performed over participant means per condition, and additionally over item means per condition. We correct for α cumulation due to multiple comparisons by Holm's method (Holm, 1979).

3.2. Results Experiment 1

The mean z-scaled ratings over participants by participants by condition are given in Table 1. The results for the statistical analysis for each of the planned comparisons were as follows:

accusative vs. illegal The difference between these two conditions is statistically significant for both modes of averaging (subjects: $p = 3.1 \times 10^{-10}$; items: $p = 7.3 \times 10^{-12}$).

accusative vs. intransitive [10] The difference between these two conditions is statistically significant for both modes of averaging (subjects: $p = 3.1 \times 10^{-10}$; items: $p = 7.3 \times 10^{-12}$).

intransitive vs. illegal The difference between both ungrammatical conditions is statistically significant (subjects: $p = 5.9 \times 10^{-4}$; items: $p = 0.016$). While both ungrammatical conditions were rated as unacceptable, ratings were lower for the illegal than for the intransitive condition.

accusative vs. dative There was no statistically significant difference, neither for subjects ($p = 0.5$), nor for items ($p = 0.87$).

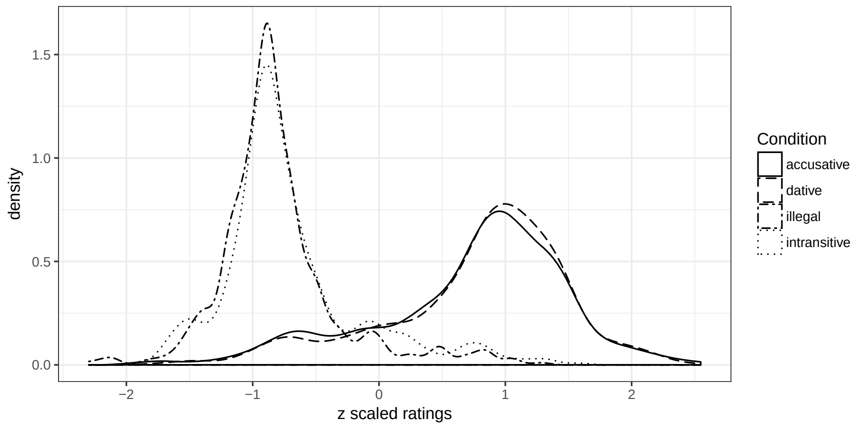


Figure 1. Density plot of z scaled magnitude estimation ratings for all conditions. We see a very clear separation between accusative and dative on one side and intransitive and illegal on the other side, and only very small differences within these pairings. We use a density plot here, that is a smoothed version of the histogram because it makes comparing conditions very convenient.

Condition	mean z-scaled ratings (SD)
accusative	0.73 (0.75)
dative	0.78 (0.72)
intransitive	-0.71 (0.56)
illegal	-0.80 (0.48)

Table 1. Mean z-scaled ratings over participants by condition for Experiment 1. Standard deviations are given in parentheses.

3.3. Discussion Experiment 1

The Magnitude Estimation study revealed that both ungrammatical conditions (intransitive and illegal) were indeed judged worse than the baseline condition. Contrary to our expectations, there was a difference between both ungrammatical conditions, with the illegal condition receiving worse ratings than the intransitive condition. We interpret this as showing that acceptability is worse when a sentence contains nonexistent words than when a sentence contains only existing words, but is ungrammatical (and probably implausible) because of a surplus argument. There was no statistically significant difference between the two grammatical conditions (accusative and dative). This suggests that even though accusative continuations were slightly preferred in the sentence completion task, the final stimulus set did not elicit different acceptability ratings for accusative and dative conditions, and that subsequent effects of accusative and dative will not reflect overall differences in acceptability. We take this to show that any potential difference between the two verb classes does not cause processing differences big enough to affect conscious ease of processing.

4. Experiment 2: Self-paced reading

In a second experiment, we monitored the on-line processing of the stimuli, using a self-paced reading paradigm. This was done to check for general quantitative differences in processing load between conditions. We expect reading times for the conditions to differ once the clause-final particle is encountered, and possibly also in the spillover-region following the sentence-final particle. We predict the following outcomes for the relevant comparisons:

- accusative-illegal: We expect reading times for illegal conditions to be slower than for accusative conditions.
- accusative-intransitive: We expect reading times for intransitive conditions to be slower than for accusative conditions.
- intransitive-illegal: Given the outcome of the Magnitude Estimation task, we would predict slightly longer reading times for particles in the illegal than in the intransitive condition.
- accusative-dative: The Lexical Reaccess hypothesis predicts enhanced lexical processing load for dative compared to accusative conditions. However, it is unclear if the differences between both conditions are strong enough to influence self-paced reading time measures.

4.1. Material and Methods

50 participants were tested. Participants were recruited via the SONA systems participant data base of the University of Konstanz. All participants spoke German as their only native language, and reported no known neurological or reading-related problems. All participants gave written informed consent. Participants received 5 Euros compensation. One participant was removed because of an error rate of 47%, considerably higher than anyone else. The remaining 49 participants were between 18 and 29 years old (mean age=23.1, SD=2.7). Eight participants were male.

Language Material The Language Material in this study is outlined in section 2. For each participant, one sentence from each of the 40 critical quartets was presented. The 40 critical quartets were interspersed with 144 filler sentences belonging to 16 different conditions, six of which were unacceptable and ten of which were acceptable. Each participant saw 184 sentences. Before the start of the experiment, participants saw five practice items.

Procedure The experiment was run in the Psycholinguistics Lab of the University of Konstanz. Stimuli were presented on a 17″ cathode ray tube monitor (Sony Trinitron Multiscan G400), connected to a Fujitsu personal computer. Stimuli were presented and reading times were recorded in Linger (Rohde, 2003). Sentences were presented word by word in a non-cumulative self-paced reading task. Critical sentences spanned a single line. After about every 46 sentences, participants were offered a break. During the course of the experiment, participants were asked to judge the acceptability of the sentence after 46 of the experimental sentences (10 of them to critical sentences). We chose to ask for acceptability judgments, rather than comprehension questions, to avoid confusing the participants with comprehension questions about ungrammatical or nonsensical sentences. Judgments were elicited with the question *Ist dieser Satz in*

Ordnung? ('Is this sentence ok?'). Participants answered the questions via presses on the Y key for 'yes' and N key for 'no'. For 30 of the 46 questions (including the questions to the critical sentences), there was an unambiguously right answer. No feedback was provided to the answers.

Data analysis 5 critical sentence quartets were removed before the final data analysis because of an error during stimulus presentation, leaving 35 sentence quartets. For the remaining sentences, reaction times longer than 6500 ms and below 150 ms were removed, because they obviously fell out of the distribution of all other data points, as judged from visual inspection. This led to the removal of 29 data points, corresponding to 0.15% of the data. After this preliminary outlier removal, we determined the best transformation of reaction times (RTs) with the aim of getting as normally distributed residuals as possible, with the secondary objective of having a simple and interpretable transformation. For this we used the `boxcox` procedure from the `R` package `MASS` (Venables & Ripley, 2002) to find the best parameter λ for the Box-Cox tranformation (Box & Cox, 1964). The resulting likelihood profile revealed a maximum close to $\lambda = -1$ which represents the inverse transformation $RT' = -1/RT$. This is a fairly common result (Baayen & Milin, 2010; Balota, Aschenbrenner, & Yap, 2013; Vasishth, Chen, Li, & Guo, 2013). The data were therefore transformed from RT (in ms) to -1/RT. The transformed data for each position were analyzed separately.

For data analysis, we fitted linear mixed effects models using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015). p-values for the models were calculated using the `ghlt` function of the `multcomp` package (Hothorn, Bretz, & Westfall, 2008). The model formula for all values of POSITION remained the same. It contains two fixed effects: The main effect of CONDITION, and the main effect of TRIAL in second order. The latter main effect was included because visual inspection revealed that the participants' reaction times became shorter in the course of the experiment. This accelaration was less pronounced during later stages of the experiment. To model this feature of the data we included the variable *Trial* in the form of a second order orthogonal polynomial to our models. This is necessary and sufficient to model the bending curves of reaction time development over time. Additionally we added random effects for SUBJECT and ITEM. For SUBJECT we included random slopes for the TRIAL-polynomial and for CONDITION. We included correlation parameters between the random intercept for SUBJECT and the two slopes for the 2nd grade polynomial, but not between the random slopes for CONDITION and anything else. The ITEM effect was restricted to a random intercept alone. For each position, we identified additional outliers by checking the residuals. We removed the data points with maximal residuals and refit, until the distribution reached optimal normality.[11] This removed 58 data points from the whole dataset, corresponding to 0.68% of the data.

Using this model, we compared the reading times for positions from the particle onward for the four planned comparisons: (1) accusative vs. illegal, (2) accusative vs. intransitive, (3) intransitive vs. illegal, (4) accusative vs. dative.

4.2. Results Experiment 2

Results are reported for single word positions, starting at the clause-final particle, and continuing into the four following positions (i.e., the spillover region). In the following, we refer to the particle position as "part", to the first post-particle word as "part+1", etc.. Only statistically significant effects are reported, unless explicitly stated otherwise. Mean reading times (retransformed to ms from $-1/RT$) at individual positions for the different conditions are given in Table 2, a visualization of the reading times is given in Figure 2.

	condition	part	part1	part2	part3	part4
1	acc	444^{+177}_{-98}	421^{+93}_{-65}	358^{+79}_{-55}	383^{+127}_{-77}	463^{+135}_{-85}
2	dat	440^{+171}_{-96}	407^{+103}_{-69}	356^{+84}_{-57}	378^{+109}_{-70}	466^{+123}_{-81}
3	intr	461^{+181}_{-102}	425^{+95}_{-65}	335^{+72}_{-51}	339^{+97}_{-62}	443^{+106}_{-71}
4	ill	465^{+170}_{-98}	430^{+114}_{-75}	337^{+76}_{-52}	333^{+93}_{-60}	432^{+99}_{-68}

Table 2. Mean reading times for Experiment 2 for all positions from the particle onward. Means and standard deviations are retransformed from $-1/RT$ to plain RT. Since the $-1/RT$-transformation is nonlinear, standard deviations are not symmetrical after retransformation.

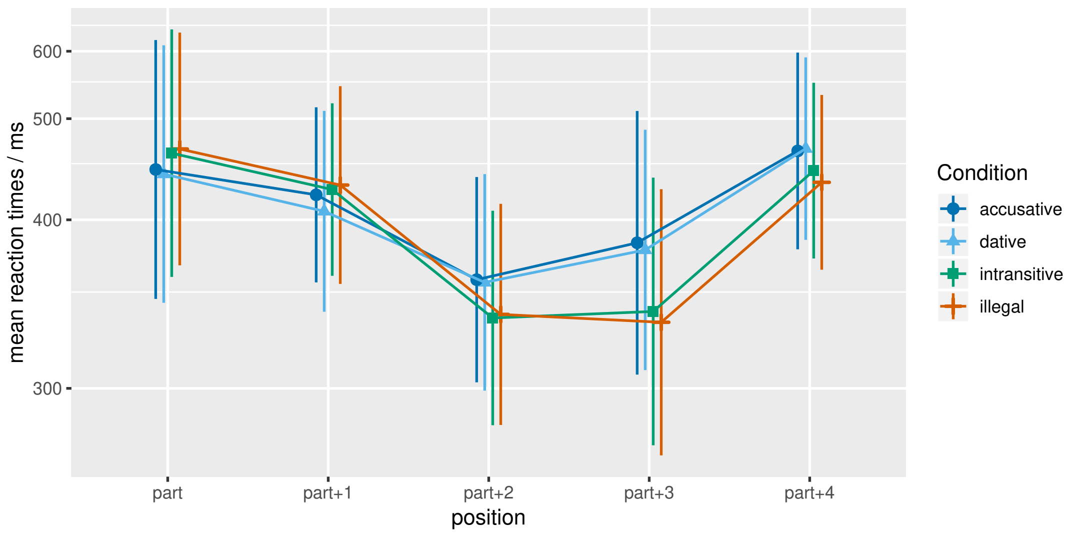


Figure 2. Selfpaced reading times for Experiment 2. Mean reading times are retransformed from -1/RT after outlier removal. The error bars depict the standard deviations.

particle There was a statistically significant difference for the comparisons accusative-illegal ($p = 0.0028$) and accusative-intransitive ($p = 0.02$). Reading times were longer in both ungrammatical conditions than in the baseline condition.

particle+1 There were no significant differences for any of the comparisons.

particle+2 There was a statistically significant difference for the comparisons accusative-illegal ($p = 1.7 \times 10^{-4}$) and accusative-intransitive ($p = 2.3 \times 10^{-4}$). Reading times were shorter in both ungrammatical conditions than in the baseline condition.

particle+3 There was a statistically significant difference for the comparisons accusative-illegal ($p \approx 0$) and accusative-intransitive ($p = 8 \times 10^{-15}$). Reading times were shorter in both ungrammatical conditions than in the baseline condition.

particle+4 There was a statistically significant difference for the comparisons accusative-illegal ($p = 4.5 \times 10^{-4}$) and accusative-intransitive ($p = 0.038$). Reading times were shorter in both ungrammatical conditions than in the baseline condition.

There were no statistically significant differences between the accusative and dative conditions, and no differences between intransitive and illegal conditions. Confidence intervals for all single comparisons are shown in Figure 3.

4.3. Discussion Experiment 2

The self-paced reading time study revealed a statistically significant difference between reading times for both ungrammatical conditions compared to the baseline. On the clause-final particle, reading times were longer for ungrammatical than for grammatical conditions. We interpret this as a reflection of the added processing cost for

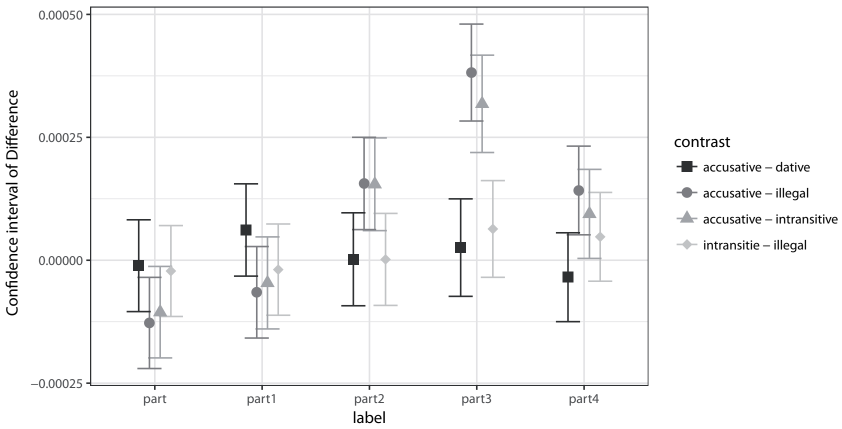


Figure 3. Confidence intervals for the performed planned comparisons. The Y axis shows the transformed reaction times $-1/RT$. In this format, a difference of $\Delta 1/RT = 0.000\,25\,1/\text{s}$ in the transformed reaction times corresponds to a ΔRT of 62.4 ms in the raw reaction times.

ungrammatical compared to grammatical conditions. This effect occurs already on the particle, i.e., the first position where the ungrammaticality of the sentences was revealed, rather than on the first spillover region. We interpret this early occurrence as an indication that the effects we measure reflect failed or difficult word recognition of nonexistent or nonmatching particle verbs, rather than in-depth syntactic parsing processes. This would also be in line with earlier findings suggesting that the combination of verb bases and particles is a lexical, rather than a syntactic process. (Cappelle et al., 2010; Hanna et al., 2017). In the spillover region (starting from the second position after the clause-final particle, and continuing until the end of the sentence), reading times were shorter for ungrammatical than for grammatical conditions. We interpret the direction of the effect as a task effect, given that we asked for acceptability judgments (instead of probe detection or comprehension questions), which would indicate that participants reject the sentence as ungrammatical once the particle has been revealed for intransitive and illegal conditions, and do not read the spillover region carefully. There was no difference between reading times for both ungrammatical conditions (intransitive and illegal), and no difference between both grammatical conditions (accusative and dative). We interpret this as an indication that there is no dramatic quantitative difference in processing cost between both grammatical structures. This implies that any potential increase in processing cost (that could be attributed to lexical reaccess to the object) is not strong enough to influence self-paced reading time measures. Any difference between both grammatical conditions that will be found in the following EEG study can therefore be assumed to reflect qualitative, rather than strong quantitative, processing differences.

5. Experiment 3: EEG

To monitor qualitative as well as quantitative processing differences between the different stimulus conditions, we used the language material outlined in section 2 in an EEG study. This technique allows us to make direct comparisons to the previous literature (Isel et al., 2005; Piai et al., 2013; Urban, 2001, 2002). In the current study, we are mainly interested in processing on the clause-final particle. Based on previous literature and the findings of Experiment 1 and Experiment 2, we make the following predictions for the planned comparisons:

- accusative-illegal: Since the clause-final particle leads to a nonexistent base-particle combination, we expect a high lexical processing load for the illegal compared to the accusative condition. Based on the findings reported by Piai et al. (2013), we expect that the unsuccessful lexicon search and the enhanced lexical processing load will be reflected in an enhanced N400 for illegal conditions. If no syntactic integration is attempted after the base-particle combination has been identified as a nonexistent word, we expect no P600 (matching the findings of Piai et al. 2013).
- accusative-intransitive: For this comparison, we expect an enhanced lexical processing load for intransitive compared to accusative conditions, reflected in an enhanced N400 for intransitive conditions. This is because the base-particle combination is unexpected in the current sentence context (and often semantically mismatching). In addition, we expect syntactic processing difficulties as the parser is unable to build a sentence structure with two argument NPs and an intransitive verb. We expect this syntactic processing difficulty to elicit an enhanced P600 for intransitive conditions.
- intransitive-illegal: Both ungrammatical conditions are expected to lead to enhanced lexical processing load reflected in an enhanced N400. If the workload of looking for a nonexistent lexical entry is measurably higher than that of looking for a lexical entry that does not match the syntactic and semantic context of the sentence, we expect the N400 to be weaker for intransitive than for illegal conditions; this would mirror the findings reported by Piai et al. (2013) for the contrast of nonexistent and semantically implausible verbs. In addition, we expect the intransitive condition to elicit a P600 due to syntactic integration difficulty. If the parser does not attempt syntactic integration of nonexistent words, we expect there to be no P600 for illegal conditions.
- accusative-dative: Following the Lexical Reaccess Hypothesis, we would predict a mildly enhanced lexical processing load for dative compared to accusative conditions, due to lexical reaccess to the object. Following Hopf et al. (2003, 1998), we expect this to be reflected in an N400. However, recent studies (Czypionka & Eulitz, 2016) suggest no strong processing differences between accusative- and dative-assigning particle verbs, and both behavioural experiments did not elicit strong dispreferences for dative compared to accusative conditions. The outcome of this comparison will allow us to reassess the Lexical Reaccess Hypothesis.

5.1. Material and Methods

Participants 26 participants were recruited via the lingexp mailing list of Konstanz LingLabs. All participants spoke German as their only native language and reported no known reading or language-related problems. Participants had normal or corrected to normal vision, were not taking any psychoactive medication and reported no neurological or psychiatric disorders. All participants were right handed, scoring 70% or higher on the Edinburgh handedness test (Oldfield, 1971). All participants gave written informed consent. For two participants, the experimental session was stopped because too many alpha waves were recorded already at early timepoints. For the participants who successfully completed the experiment, two participants were removed because of poor data quality (over 25% of data removed). The mean age of the remaining 22 participants (12 male) was 24.7 years (SD = 3.4). Participants received 30 Euros compensation.

Procedure The participants were seated in a comfortable chair in front of a computer screen, with an average distance of about 180 cm, in an electrically shielded EEG recording chamber. The experiment consisted of an instruction phase and the experimental phase. Participants were first instructed orally and then again in written form on the screen during the instruction phase. Words were presented visually in the center of a computer screen using the Presentation software by Neurobehavioral Systems Inc. (version 16.1), in white 40 pt Arial font on a black screen. During the experiment, participants held a two-button response box in their hands. Participants answered the questions by pressing the left or right response button, respectively.

Stimulus presentation All participants saw all items in all conditions, i.e. $40 \times 4 = 160$ critical sentences. To avoid repetition effects, the critical items were interspersed with $40 \times 4 \times 2 = 320$ filler sentences. Half of the filler sentences were ungrammatical, half of the filler sentences were grammatical. The stimulus list was split in two blocks to avoid tiring out the participants. Each block contained two sentences from each critical sentence quartet, one grammatical and one ungrammatical, to even out the distribution of verbs and base verbs. Half of the participants saw the first block first, the other half saw the second block first. Testing one block took about 40 minutes. There was a pause between blocks. Participants were offered the chance to take a longer break between the two blocks, stand up, move, drink or eat. In addition, there were three short breaks in each block, occurring after about 80 sentences. If necessary, some electrodes were re-prepared during the breaks.

Sentence presentation started with an asterisk in the center of the screen, presented for 500 ms, followed by a 200 ms blank screen. The sentences were segmented into multiple-word and single-word chunks, presented one after the other. The pattern was: [*Lehrerinnen* | *hören* | *Schulbands* | *auf Konzerten* | *geduldig* | ***an*** | *wenn* | *sie eingeladen sind.*]

All chunks but the last were presented for 800 ms, the last chunk was presented for 900 ms. Between chunks, a blank screen was presented for 200 ms. After 96 of the sentences, we asked the participants to decide whether the sentence was grammatical or not. The question always was *Ist dieser Satz grammatisch?* ('Is this sentence grammatical?'). [12] Participants had to press buttons on a button box to answer with either 'yes' or 'no'. The buttons corresponding to 'yes' or 'no' were randomly assigned for each question. Below the question, the words "JA" and "NEIN" ('yes' and 'no') were presented on the screen, fitting the orientation assigned to the answers for this specific question (i.e., whether the 'yes'-button was on the right or on the left). We switched answer orientations to provide the participants with a motivation to stay alert, and, more importantly, to avoid the buildup of LRPs (lateralized readiness potentials, see Coles 1989; Kornhuber and Deecke 1965; Shibasaki and Hallett 2006). 32 of the sentences with questions were sentences in critical conditions, with 8 of each critical condition.

The right answer to the critical questions was 'yes' in 16 cases and 'no' in 16 cases. 64 of the sentences with questions were filler sentences, for 40 of which the right answer was 'yes' and for 24 of which the right answer was 'no'.

EEG recording The EEG was recorded with 61 Ag/AgCl sintered ring electrodes attached to an elastic cap (EasyCap, Herrsching) and connected to an Easy-Cap Electrode Input Box (EiB32). Electrodes were positioned in the equidistant 61-channel arrangement provided by EasyCap (see `http://easycap.brainproducts.com/e/electrodes/13_M10.htm` for electrode layout). The EEG signal was amplified with a BrainAmp DC amplifier with a bandpass of 0.016 Hz - 250 Hz (Brain Products, Gilching) connected to a computer outside of the EEG chamber (via USB2 Adapter, Brain Products, Gilching). The signal was recorded with a digitization rate of 500 Hz (Brain Vision Recorder, Brain Products, Gilching). Eye movements were monitored by recording the electrooculogram (IO1, IO2, Nz). The ground electrode was located on the right cheek.

Data processing Data were processed using the Brain Vision Analyzer 2 software (Brain Products, Gilching). Raw data were inspected visually. Time windows including strong, visible artifacts and breaks were manually removed. Next, an ICA blink correction was performed for the remaining data, using the Slope Algorithm for blink detection. After the blink correction, data were again inspected visually to monitor successful blink correction. A spline interpolation was performed for channels that showed long stretches of noisy data. Interpolation was only performed for electrodes with at least three surrounding non-interpolated electrodes (two for electrodes on the outermost ring). After the interpolation, all electrodes were re-referenced to average reference. An Automatic Raw Data Inspection was performed for the re-referenced data (maximal allowed voltage step: 50 µV / ms; maximal allowed difference: 100 µV / 200 ms; minimal/maximal allowed amplitudes 200 µV / -200 µV; lowest allowed activity: 0.5 µV / 100 ms). Before segmentation, the remaining raw data were filtered with Butterworth zero phase bandpass filters. The low cutoff frequency was 0.05 Hz (12 dB / oct), the high cutoff frequency was 70 Hz, (12 dB/oct). After filtering, data were segmented into time windows time-locked to the onset of the clause-final particle. Time windows began at -200 ms before the onset of the particle, and ended at 1000 ms after the onset of the particle. A baseline correction was performed for the 200 ms before the onset of the particle. Averages were calculated per participant for all four conditions. Participants with less than 30 trials in one of the four conditions were excluded from data analysis, leading to the exclusion of four participants (see *Participants* subsection above). For the remaining 22 participants, on average 8.2 % of the data were rejected (SD = 5.8%), so that all condition means were calculated from 40 to 30 segments. The mean number of segments per condition were: accusative = 37 (SD = 2), dative = 37 (SD = 2), intransitive = 37 (SD = 2), illegal = 37 (SD = 3). For data presentation, Grand Averages were smoothed with an additional 10 Hz low-pass filter.

Parametrization and statistical testing. Time windows were chosen based on our expectations outlined above and on visual inspection of the data, with reference to previous studies. For checking our predictions, it was central to assess differences between conditions in two time windows. We expected differences in the N400 time window (based on Isel et al. 2005; Piai et al. 2013; Urban 2001, 2002 for accusative compared to illegal and intransitive base-particle combinations, and Hopf et al. 2003, 1998 for accusative compared to dative base-particle combinations). In addition, we wished to assess whether illegal and intransitive combinations led to enhanced lexical processing cost or reanalysis (which we expected to become visible in an enhanced P600 or other later components usually associated with syntactic processing or lexical-semantic integration difficulties of verbs and potential arguments, Hoeks et al. 2004; Kim and Osterhout 2005; Kuperberg et al. 2007). The chosen time windows were 370 to 470 ms after the onset of the clause-final particle for the N400 time window and 600 to 800 ms after the onset of the clause-final particle for the P600 time window. The N400 time window was chosen by identifying a clearly visible peak in the data and extending the time window symmetrically around this peak. The P600 time window was chosen by visually identifying the time point when a strong positive deflection started to become visible for illegal and intransitive conditions compared to the accusative condition around 600 ms (fitting the starting point of the P600 time window reported by Hoeks et al. 2004; Kim and Osterhout 2005, and slightly later than the one reported by Kuperberg et al. 2007). This positive deflection remained for the rest of the word presentation time.

We modeled the mean voltages for the two time windows by generalised additive mixed models (GAMMs, Baayen, Vasishth, Bates, and Kliegl (2016); Kryuchkova, Tucker, Wurm, and Baayen (2012); Tremblay and Baayen (2010); Wood (2006); Zuur, Ieno, Walker, Saveliev, and Smith (2009)) in R, using the `mgcv` (Wood, 2011) and `itsadug` (van Rij, Wieling, Baayen, & van Rijn, 2017) packages.

To allow an easy comparison of our results with the literature, we also report the results of a more traditional statistical analysis in Appendix 1, using a subset of electrodes grouped into regions of interest, and reporting the outcomes of repeated-measures ANOVAs for mean amplitudes in these ROIs. Since our statistical testing approach is not yet widespread, we will outline the advantages of GAMMs over ANOVAs, and the development of the underlying model in some detail before reporting the results. The outcome of the final model is given in the results section. Further details of the model are additionally summarised in Appendix (2).

GAMMs model nonlinear dependencies of response variables on continuous predictors by nonlinear smoother functions. These smoother functions can be specified to bend as flexibly and wiggly as is justified by the data. Smoothers are not restricted to one variable, but can depend on two or more continuous variables[13]. We use this feature in order to account for spatial relations between individual electrodes: We parametrised the electrode positions in a two-dimensional plane defined by two spatial variables, X and Y.[14]We do not parametrise the electrode positions on the scalp in three dimensions, since the problem is in essence two-dimensional, just like a globe can be projected onto a two-dimensional map without information loss. That is of course exactly the way EEG data are usually plotted. This approach contrasts favourably with a classical ANOVA analysis, where all electrodes within the same ROI are considered equal and equally informative, irrespective of whether they are situated on the boundary or in the middle of the ROI, and regardless of their relative positions. Another, closely related advantage of the GAMM approach is that it removes the necessity to decide on a specific parametrization or partition into ROIs, or else to select single exemplary electrodes from the full electrode set that are to be analyzed in greater detail. Those decisions always remain somewhat arbitrary, however carefully justified. Using the GAMM approach, interpreting the data becomes easier because there is no such selection and partition choice that could artificially modulate effects. All electrodes enter into the model on an equal footing, and the results directly show the spatial distribution of the effects.

To detect the impact of our experimental condition, we modeled the mean voltage amplitude at all electrodes by two smoothers in X and Y, one for one of the compared conditions, the other for the difference between this and the second of the compared conditions. This difference smoother is the main tool of analysis because it informs us about the existence and location of significant effects. To account for the overall absolute potentials, condition was added as a simple fixed effect as well. Systematic differences between subjects are accounted for by random smoothers, equally two-dimensional in X and Y.

The residuals of GAMM models constructed as described so far showed heavy spatial correlations. Residuals from close electrodes were positively correlated, while residuals from opposing sides of the scalp correlated negatively. This feature could tentatively be explained by the geometry of the dipoles contributing to the measured potentials whose gradients form outcoming and incoming field lines on opposing sides in direction of their axes. No matter how these correlations exactly arise, they invalidate the model assumption that the residuals are independent of each other. The consequence of this might be anti-conservative p-values (Baayen, van Rij, de Cat, & Wood, 2016). That problem has to be dealt with. We extended the model to incorporate this behaviour in the following way. Let us consider one particular electrode E_i at some fixed combination of participant and condition. The aim is to explicitly account for the influence of the potentials at all other electrode positions E_j. For this we summarise those potentials into one single number, which can be fed into the model as a predictor. While that number will be some kind of average, the standard average, the arithmetical mean, does not work, because the influence of an electrode E_j on E_i will depend on the distance d_{ij} between the two. This situation implies a weighted mean. The weights are a function of the spatial distance d_{ij} of E_i and E_j. Since the weights are meant to model a three-dimensional effect (i.e., correlation between directly neighbouring electrodes, and electrodes from opposing sides of the skull), we used three-dimensional electrode coordinates to compute this distance (Table 1 in 2). More specifically, the weight function we used decays quadratically with the distance d_{ij} of electrodes E_i and E_j. This specific shape was used for two complementary reasons. First, it can be theoretically motivated, given that the potential of an electric dipole decays quadratically. Second, we empirically tested other decay schemes with considerably worse results, as judged by AIC (Akaike, 1974). The weighted average defined that way enters the model as fixed effect. The whole procedure can be seen as a substantial extension of the idea proposed in Baayen and Milin (2010) for the same problem in one (temporal) dimension.

We provide the information about which electrodes contribute to statistically significant effects in a graphical format. In our visualizations, we color-code the areas where the confidence bands around the difference smoothers do not include 0. This means that colored electrodes and areas are the regions where statistically significant differenes between the respective conditions arise. This depiction is both more intuitive and more precise than the traditional practice of reporting the interaction of ROI and CONDITION effects.

5.2. Results Experiment 3

The mean error rate for the participants included in the study was 12% (SD = 12) for questions regarding the critical conditions. Mean voltage difference maps and EEG curves for selected electrodes are presented in Figure 4. Visualizations of the statistical significances of reported effects are given in Figure 5 for the N400 time window, and Figure 6 for the P600 time window.

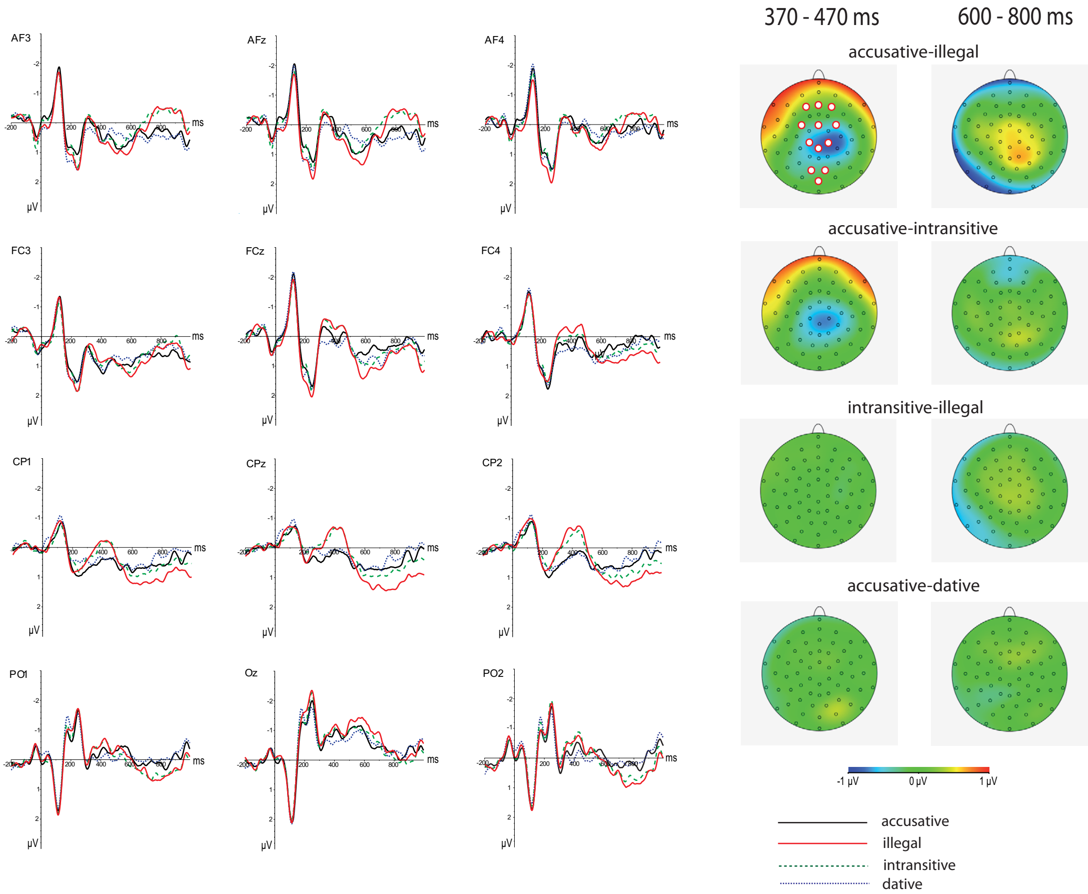


Figure 4. Experiment 3: Grand average ERPs for selected electrode sites and difference maps are shown. Mean voltage difference maps for the planned comparisons are given on the right side for both time windows. The electrodes selected for illustration are marked in the maps. Negativity is plotted upwards.

Outcomes for the planned comparisons in the 370–470 ms time window were as follows:

accusative-illegal: The difference between accusative and illegal conditions was statistically significant (estimated degrees of freedom $edf = 12.5$, $p = 2 \times 10^{-5}$). Overall significance stems from the potentials being more negative for illegal than for accusative conditions at central electrode sites, and less negative at lateral posterior locations, see Figure 5a.

accusative-intransitive: The difference between accusative and intransitive conditions was statistically significant ($edf = 11$,p=0.002). Overall significance stems from the potentials being more negative for intransitive than for accusative conditions at central electrode sites, and less negative at lateral posterior electrode sites, see Figure 5b.

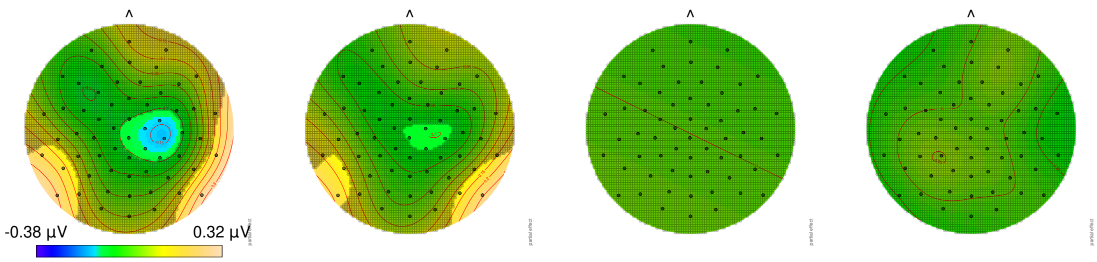
intransitive-illegal: The difference between intransitive and illegal conditions did not reach statistical significance. Curves for both conditions ran parallel, and were more negative-going than for the accusative baseline.

accusative-dative: There was no statistically significant difference between accusative and dative conditions in this time window. Waveforms for both conditions ran closely parallel.

As visualised in Figure 5, the difference smoother deviates significantly from a simple plane around zero for the comparisons accusative-illegal and accusative-intransitive, while the results of the other comparisons remain insignificant. The patterns are similar within this time window. The main contribution to the difference comes from the illegal and intransitive conditions showing more negative mean voltages than the accusative condition at central electrode sites. Descriptively, this difference is more pronounced in the accusative-illegal comparison than in the accusative-intransitive comparison; however, this difference did not reach statistical significance. Additional contributions come from lateral-posterior electrode sites on both sides of the skull. We interpret the negativities found in the 370 to 470 ms time window for illegal and intransitive conditions relative to the accusative baseline as an N400 effect.[15]

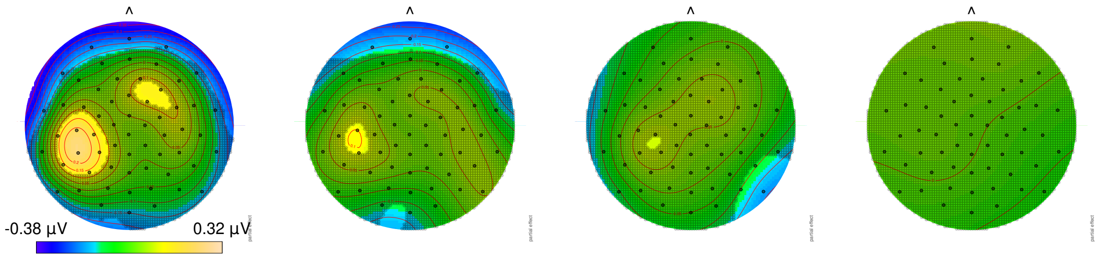
Outcomes for the planned comparisons in the 600 to 800 ms time window were as follows:

accusative-illegal: The difference between accusative and illegal conditions was sta-



(a) illegal-accusative (b) intransitive-accusative (c) illegal-intransitive (d) dative-accusative

Figure 5. The GAMM results for the 370 to 470 ms time window. The individual panels represent the outcomes for the different planned comparisons. We show the model fit for the two-dimensional difference smoother. Within the greyed out areas the 0-potential plane lies within the confidence bands. Brightly coloured regions indicate the regions responsible for overall significance.



(a) illegal-accusative (b) intransitive-accusative (c) illegal-intransitive (d) dative-accusative

Figure 6. The GAMM results for the 600 to 800 ms time window. The individual panels represent the outcomes for the different planned comparisons. We show the model fit for the two-dimensional difference smoother. Within the greyed out areas the 0-potential plane lies within the confidence bands. Brightly coloured regions indicate the regions responsible for overall significance.

tistically significant ($edf = 17.3$, $p = 4 \times 10^{-8}$). Overall significance stems from waveforms being more positive for illegal than for accusative conditions at left-central sites and being more negative at frontal positions, see Figure 6a. Descriptively speaking, this positivity was visible at central, left-central and right-central sites.

accusative-intransitive: The difference between accusative and intransitive conditions was statistically significant ($edf = 17.3$, $p = 0.002$). Overall significance stems from waveforms being slightly more positive for intransitive than for accusative conditions at left-central sites and more negative at the right frontal positions, see Figure 6b. Descriptively speaking, this positivity was visible at central, left-central and right-central sites.

intransitive-illegal: The difference between intransitive and illegal conditions was statistically significant ($edf = 9.94$, $p = 0.02$). Waveforms for both conditions ran roughly parallel, but were slightly more positive-going for illegal than intransitive conditions at left-central sites, and more negative-going at the most frontal electrode positions.

accusative-dative: There was no statistically significant difference in this time window.

As visualised in Figure 6, the difference smoother deviates significantly from a simple plane around zero for the comparisons accusative-illegal and accusative-intransitive. The main contributions to this effect come from two regions. At medial electrode sites, mean voltages are more positive for illegal and intransitive conditions than for the accusative condition. At frontal sites, mean voltages are more negative for illegal and intransitive conditions than for the accusative condition. [16]

We interpret the positivities found in the 600 to 800 ms time window for illegal and intransitive conditions relative to the accusative baseline condition as a P600 effect. The small, but statistically significant difference between intransitive and illegal conditions supports the visual impression that the P600 is more enhanced in illegal than in intransitive conditions.

5.3. Discussion Experiment 3

We will discuss the outcomes of the planned comparisons separately, before addressing a comparison to the outcomes of Experiments 1 and 2 in the General Discussion.

accusative-illegal Our predictions led us to expect higher lexical workload for illegal compared to accusative conditions on the clause-final particle, reflected in an enhanced N400 for illegal conditions. This expectation was confirmed. In addition, there was an enhanced P600 for illegal compared to accusative conditions, indicating increased syntactic processing difficulty. While the enhanced N400 matches findings reported for Dutch by Piai et al. (2013) (for German also by Urban 2001, 2002, and, using a different stimulus design, Isel et al. 2005), the P600 was not reported before.

accusative-intransitive Our predictions led us to expect an enhanced N400 for intransitive compared to accusative conditions on the clause-final particle. While the base-particle combinations in the intransitive condition do exist, the particle leading to an intransitive verb is unexpected in these sentence contexts, and should be difficult to retrieve from the lexicon for this reason. This expectation was confirmed. In addition, we expected an enhanced P600 for intransitive compared to accusative conditions, since we assumed that an existing verb with a non-matching subcategorization frame would elicit syntactic processing difficulty. This expectation was confirmed.

intransitive-illegal This comparison was performed to assess if the violations caused by the two ungrammatical conditions elicited different types of processing difficulties. While descriptively, the N400 and P600 look more pronounced for illegal than for intransitive conditions, this difference was only statistically significant in the P600 time window. While this difference is visible and reaches statistical significance, it is less pronounced than the differences for both ungrammatical conditions relative to the accusative baseline. The fact that we did not find a difference in the N400 time window does not mirror the findings reported by Piai et al. (2013) for their second experiment, where the N400 elicited by nonexistent base-particle combinations was significantly stronger than the one elicited by implausible base-particle combinations. This difference can be explained with differences in the stimulus material used in both studies: While the implausible condition in the earlier study did not introduce a grammatical violation, the intransitive condition in the current study systematically introduced a subcategorization violation on top of any lexical and semantic processing difficulty. We assume that this makes the latter a more severe violation than the former, thereby eliciting components almost as strong as those elicited by nonexistent words in the illegal condition.

To sum up the comparison of both ungrammatical conditions (illegal and intransitive) to the baseline (accusative), both ungrammatical conditions elicited an N400, followed by a P600, on the clause-final particle. Based on the literature (Isel et al., 2005; Piai et al., 2013; Urban, 2001, 2002), the N400 was expected, given that lexical processing load should be enhanced both for verbs that do not exist and for verbs that do not match the syntactic and semantic context of the sentence. The P600 was expected for the intransitive condition, given that it systematically introduces a syntactic violation. However, earlier comparable studies (Isel et al., 2005; Piai et al., 2013) did not report a

P600 for nonexistent base-particle combinations. Nonetheless, the fact that we find an N400-P600 sequence fits in well with the account proposed by Brouwer et al. (2012). Following this account, we take the N400 to reflect increased processing load for lexical access - ultimately unsuccessful in the case of illegal particle verbs, and not exactly facilitated by the preceding transitive sentence environment in the case of intransitive verbs. The P600 can be taken to reflect increased processing load for integration of these words (or nonwords) into the sentence context.

How can we explain the difference between our own findings and those of earlier studies? The difference to the findings reported by Isel et al. (2005) are not too surprising, given that in this study, nonexistent particle verbs were constructed in a very different way from our own (i.e., using verbs as bases that never are used as bases, in contrast to combining potential bases with illegal particles). The difference to the findings reported by Piai et al. (2013) is surprising at first sight. However, a closer look at the stimuli employed in their second experiment makes this difference more plausible. In the current study, all sentences followed the same pattern (outlined in section 2), and the particles were always in identical positions in the sentence in all conditions. In the study by Piai et al. (2013), the stimulus set for the second experiment contained sentences with different structures, so that the particles (that the EEG was taken from) sometimes occurred directly after the base verb (*Spreeuwen strijken(V) neer(P) in Malden en veroorzaken problemen.* ‘Starlings alight in Malden and cause problems.’), sometimes with more constituents occurring between the two (*De staat werpt(V) het huidige regime omver(P) en vervangt het door ...* ‘The state overthrows the current regime and substitutes it by. . . ’). Moreover, the existing particle verbs used in this earlier study were not matched for the number of arguments (*Ik draaide(V) Sanne de rug toe(P) omdat ik boos was.* ‘I turned my back to Sanne because I was angry’), so that it is very likely that the baseline reflects many different types of successful syntactic integration. This may have blurred a potential contrast between the grammatical condition and the one with nonexistent verbs in this earlier study. It is thus unlikely that the lack of a P600 in the results reported by Piai et al. (2013) means that the parse is abandoned immediately upon encountering the particle that leads to a nonexistent base-particle combination.

accusative-dative Following the Lexical Reaccess Hypothesis, we would have expected a mildly enhanced N400 for dative compared to accusative conditions, reflecting lexical reaccess to the object to check for morphological licensing of dative (see Hopf et al. 2003, 1998). This N400 should have been weaker than the one elicited by either of the ungrammatical conditions, since dative conditions were not ungrammatical or nonsensical, and the sentence completion task and ME study show that dative continuations are not unexpected and do not lead to reduced acceptability. However, based on earlier findings comparing accusative- and dative-assigning particle verbs that were presented as a whole at the clause-final position (Czypionka & Eulitz, 2016, to appear), we would have expected no enhanced N400 for dative compared to accusative conditions. Our findings reveal no indications of enhanced N400 or P600 for dative compared to accusative conditions. Neither does visual inspection of the data suggest a slight enhancement of the N400 for datives. If lexical case marking should really have caused lexical reaccess to the object, as proposed by the Lexical Reaccess Hypothesis, we would have expected an enhanced N400 (following Hopf et al., 2003, 1998). The fact that we did find an N400 for accusative-illegal and accusative-intransitive comparisons shows that our experimental setup is in principle well-suited to find N400 effects elicited by different types of enhanced processing load. This makes it unlikely that there is a mild N400 effect in the accusative-dative comparison that we missed for technical reasons. While accusative- and dative-assigning particle verbs may possibly differ in processing (for example, due to semantic differences, see Meinunger 2007; Svenonius 2010), our findings suggest that these differences do not include a measurably enhanced lexical processing load for datives due to lexical reaccess to the object.

6. General Discussion and Conclusion

The goal of the present paper was to shed light on processes of lexical, syntactic and semantic integration happening on the clause-final particles of split base-particle combinations. In contrast to earlier studies, our study had a special focus on syntactic changes introduced by the clause-final particle. In our stimuli, we compared sentences that had identical base verbs in V2, but were completed at the clause final position by a particle leading to accusative-assigning two-place verbs, nonexistent base-particle combinations, intransitive one-place verbs, and dative-assigning two-place verbs. The processing of these stimuli was monitored in three studies: Acceptability ratings were used to check subtle acceptability differences between grammatical and ungrammatical conditions, and to check our own intuitions. Self-paced reading times were used to check overall quantitative differences in workload between the different conditions. Finally, EEG measurements were used to check quantitative as well as qualitative processing differences between conditions, and to allow a comparison to the literature.

The acceptability rating study showed worse ratings for both ungrammatical conditions (illegal and intransitive) than for both grammatical conditions (accusative and dative). The illegal condition received slightly worse ratings than the intransitive condition, suggesting that in offline acceptability ratings, participants perceive the violation introduced by a nonexistent word as stronger than the one introduced by a word that does not match the syntactic frame of the sentence. There was no statistically significant difference in ratings between accusative and dative conditions, suggesting that both conditions were equally acceptable and that any effect found in subsequent studies could not be reduced to differences in acceptability.

The self-paced reading times study again revealed differences between the ungrammatical conditions and the grammatical conditions, visible in longer reading times at the particle and shorter reading times at subsequent positions for ungrammatical compared to grammatical conditions (the latter likely a task effect, since the grammaticality judgment could be made following the particle in the ungrammatical conditions). Unlike the results of the acceptability rating study, the results of the self-paced reading time study did not reveal subtle distinctions between processing load for both ungrammatical conditions, suggesting that this online processing measure is not sensitive to the preferences visbile in the offline acceptability rating study. Neither was there a statistically significant difference between accusative and dative conditions in this study, suggesting that the overall processing load for both conditions is comparable, and that possible distinctions visible in the later EEG experiment would reflect qualitative rather than quantitative differences in processing.

The EEG study revealed an N400 effect for both ungrammatical conditions compared to the accusative baseline condition. There was no statistically significant difference between the dative and the accusative condition in any of the three experiments. Particles leading to dative-assigning particle verbs did not elicit an enhanced N400 relative to the accusative condition. This is in contrast to the predictions we formulated based on the Lexical Reaccess Hypothesis proposed by Bayer et al. (2001), which holds that upon encountering a dative-assigning verb in the absence of previous overt morphological marking, the lexical entry of the object has to be reaccessed to check for dative morphology (leading to an enhanced N400). The fact that an N400 effect was visible for both ungrammatical conditions makes it likely that our experimental setup is in principle well-suited to find N400 effects.[17] Importantly, this finding does not preclude that lexical case marking verbs (i.e., dative-assigning two-place verbs) are processed differently from structural case marking verbs (i.e., accusative-assigning two-place verbs) in general. However, it makes it unlikely that increased processing load caused by lexical reaccess to the object is at the root of these processing differences, since these should have been visible in our findings. Although the processing of dative does not seem to cause increased lexical processing load in our experiments, our findings do not imply that dative assigned by two-place particle verbs should not be considered a lexical case (as illustrated, perhaps trivially, by the observation that dative assigned by particle verbs is retained under passivization, while accusative assigned by particle verbs is not). However, this finding suggests that the revelation that a case-ambiguous object bears lexical case does not necessarily translate to measurably enhanced lexical processing cost. Currently, there is an ongoing debate in the theoretical literature about the argument status of the objects of particle verbs, i.e., whether they should be considered the objects of the base verb, the particle (which would work comparable to an incorporated preposition), or a base-particle complex (see McIntyre 2007, 2015 for extensive overviews over the different proposals). If a theoretical proposal about this issue were formulated in a way that allows clear predictions for processing, these predictions could be put to the test in future experiments on the processing of existing particle verbs in grammatical sentences. This would be necessary to allow a conclusive interpretation of our negative finding. So far, our findings for the accusative-dative comparison encourage us to reassess the role of lexical case marking in the current sentence comprehension literature. Proposals for explaining the processing differences between accusative- and dative-assigning two-place verbs have assumed that lexical differences as proposed by the LRH (Bayer et al., 2001) or thematic-semantic differences as proposed by the extended Argument Dependency Model (eADM, Bornkessel-Schlesewsky and Schlesewsky 2006, 2009, 2013) are at the root of measurable lexical case marking effects. Both types of explanations should lead us to predict processing differences between particle verbs in the accusative and dative conditions.

Thus, our findings do not match the predictions of the LRH, and neither the predictions of the eADM. However, the earlier findings that these proposals are based on were elicited using a mix of simple verbs and particle verbs assigning accusative and dative, which differ from each other in many dimensions. The only study explicitly testing simple and particle verbs separately failed to find case marking effects for non-split particle verbs, while simple verbs showed clear case-marking effects (Czypionka & Eulitz, 2016, to appear). We therefore take our negative result from the accusative-dative comparison as another hint that the contrast between lexical and structural case marking does not play identical roles in the processing of simple and particle verbs. The processing of lexical case thus merits further attention, both from a theoretical perspective and to allow an incorporation of its role in models of sentence comprehension.

The contrast between the ungrammatical conditions and the accusative baseline allows for an easier interpretation. While descriptively, the N400 effect for the illegal condition looked slightly more pronounced than for the intransitive condition, the difference between intransitive and illegal conditions did not reach statistical significance. The N400 effect was expected for both ungrammatical conditions, given the previous literature (Isel et al., 2005; Piai et al., 2013; Urban, 2001, 2002). In addition, there was a P600 effect for both ungrammatical conditions compared to the accusative baseline condition. This P600 effect has not been reported before for nonexistent base-particle combinations (corresponding to our illegal condition); we assume that it emerged in the current study because the stimuli were kept rigorously parallel between conditions up to the point of the clause-final particle (see the discussion of Experiment 3 for a detailed comparison to previous studies).

Taken together, our findings for illegal and intransitive conditions indicate an enhanced lexical processing load upon encountering a clause-final particle that combines to a base-particle combination without a lexical entry (i.e., that does not exist), or that is unexpected (because it does not fit the syntactic structure of the preceding sentence, not having an object role for one of the preceding DPs; as well as not making semantic sense in many instances). In addition, these particles lead to increased difficulty in syntactic integration, reflected in an enhanced P600. This latter finding suggests that the parse is not abandoned immediately upon encountering a particle leading to a nonexistent base-particle combination. The results fit well with current accounts of the underlying processes reflected in N400 and P600 effects, in this instance particularly with the lexical access - syntactic integration account proposed by Brouwer et al. (2012). The enhanced P600 for ungrammatical conditions could reflect two different types of syntactic combination. The first possibility is that it reflects general difficulties during the build-up of sentence structure (because the sentence turns out to not contain an existing verb in the illegal condition, or because it turns out to contain an NP without an argument role in the intransitive condition). The second possibility is that it reflects difficulties stemming from the resolution of the lexical and syntactic dependency between the base and the particle. We favour the first explanation, mainly because earlier studies indicate that particle verbs are accessed as lexical units and are not combined via syntactic processes, both in isolation (Cappelle et al., 2010) and in sentence context (Hanna et al., 2017).

In sum, our findings have contributed to a clearer picture of the processing of separated particle verbs in German sentence comprehension. We replicate and extend earlier findings indicating lexical processing difficulty for ungrammatical conditions, but not for grammatical conditions that could be expected to show enhanced lexical processing cost based on the current sentence comprehension literature. In addition, we are able to show that carefully constructed ungrammatical conditions can also elicit reflections of syntactic integration difficulty that were previously not reported. Our findings show that both lexical and syntactic processes play a role in the comprehension of German main clauses with split particle verbs. These findings can inform future studies aiming to tackle the unique interplay between lexical and syntactic combination processes in this type of sentence.

7. Acknowledgements

The authors gratefully acknowledge the contributions of Oleksy Bobrov (technical support), Alica Abberger, Christina Gozebina and Marc Meisezahl (student assistant support during data recording and evaluation). For helpful discussions, we are grateful to Nicole Dehé, Josef Bayer and the members of the Konstanz syntax colloquium 2014-2016. Additional thanks are due to Colin Phillips and the participants of the workshop *A special workshop on psycholinguistics with Prof. Colin Phillips* (December 2015, Center for Experimental Research on Natural Language, University of Wrocław, Poland), and to Jacolien van Rij and the participants of the workshop *Statistical analysis of linguistic data* (April 2017, University of Konstanz). This work was supported by the Foundation for Polish Science under Grant number F5/09/P/2013 of January 27, 2014 (Decomposing Categories in the Brain: Focus on linguistic categories expressing eventualities, subproject 3: Processing of discontinuous verbs in German. The role of particles for Case licensing). This work was furthermore supported by grants of the University of Konstanz and and by the German Research Foundation (DFG) under Grant number FOR2111 (Questions at the interfaces, TP1: Die Funktion deutscher Modalpartikel in Fragen).

8. Author Note

Correspondence concerning this article should be addressed to Anna Czypionka, Department of Linguistics, University of Konstanz, Fach D 185, 78457 Konstanz, Germany. Email: anna.czypionka@uni-konstanz.de

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. Retrieved from http://www.ualberta.ca/~baayen/publications/BaayenMilin2010.pdf

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2016, January). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. *ArXiv e-prints*. (https://arxiv.org/abs/1601.02043)

Baayen, R. H., Vasishth, S., Bates, D., & Kliegl, R. (2016). The cave of shadows. addressing the human factor with generalized additive mixed models. *ArXiv e-prints*. (https://arxiv.org/abs/1511.03120)

Bader, M. (1996). *Sprachverstehen. Syntax und Prosodie beim Lesen*. Opladen: Westdeutscher Verlag.

Bader, M. (2012). The German *bekommen* passive: A case study on frequency and grammaticality. *Linguistische Berichte*, *231*, 249–298.

Bader, M., Meng, M., & Bayer, J. (2000). Case and reanalysis. *Journal of Psycholinguistic Research*, *29*(1), 37–52.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1563–1571.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bayer, J., Bader, M., & Meng, M. (2001). Morphological underspecification meets oblique case: Syntactic and processing effects in German. *Lingua*, *111*, 465–514.

Blume, K. (2000). *Markierte Valenzen im Sprachvergleich*. Tübingen: Max Niemeyer Verlag.

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-Schlesewsky, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787– 821.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, *3*(1), 19–58.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*(2), 211—252.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143.

Cappelle, B., Shtyrov, Y., & Pulvermüller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, *115*(3), 189–201.

Coles, M. G. (1989). Modern mind-brain reading: psychophysiology, physiology, and cognition. *Psychophysiology*, *26*(3), 251–269.

Czepluch, H. (1996). *Kasus im Deutschen und Englischen. Ein Beitrag zur Theorie des abstrakten Kasus*. Tübingen: Narr.

Czypionka, A. (2014). *The interplay of object animacy and verb class in representation building* (Doctoral dissertation, Humboldt-Universität Berlin). Retrieved from http://edoc.hu-berlin.de/dissertationen/czypionka-anna-2013-12-05/PDF/czypionka.pdf

Czypionka, A., & Eulitz, C. (2016). *Case marking affects the processing of animacy with simple verbs, but not particle verbs: An event-related potential study.* Konstanz, Germany.

(presented at *The syntax of argument structure: empirical advancements and theoretical relevance.* 38th Annual Meeting of the DGfS, organized by Alexiadou, Artemis and Verhoeven, Elisabeth)

Czypionka, A., & Eulitz, C. (to appear). Lexical case marking affects the processing of animacy in simple verbs, but not particle verbs: Evidence from event-related potentials. *Glossa*.

Czypionka, A., Spalek, K., Wartenburger, I., & Krifka, M. (2017). On the interplay of object animacy and verb type during sentence comprehension in German: ERP evidence from the processing of transitive dative and accusative constructions. *Linguistics*, *55*(6).

Dehé, N. (2015). Particle verbs in Germanic. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-Formation: An International Handbook of the Languages of Europe, vol.1* (p. 611-626). Berlin / Boston: Mouton de Gruyter.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547-619.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192.

Grimm, S. (2010). Semantics of case. *Morphology*, *21*(3-4), 515–544.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483.

Haider, H. (1993). *Deutsche Syntax - generativ.* Tübingen: Narr.

Haider, H. (2010). *The Syntax of German.* New York: Cambridge University Press.

Hanna, J., Cappelle, B., & Pulvermüller, F. (2017). Spread the word: MMN brain response reveals whole-form access of discontinuous particle verbs. *Brain and Language*, *175*, 86–98.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., & Geyken, A. (2011). dlexDB — eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *32*(4), 10–20.

Hoeks, J., Stowe, L., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59–73.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*(6), 65–70.

Hopf, J.-M., Bader, M., Meng, M., & Bayer, J. (2003). Is human sentence parsing serial or parallel? evidence from event-related brain potentials. *Cognitive Brain Research*, *15*(2), 165–177.

Hopf, J.-M., Bayer, J., Bader, M., & Meng, M. (1998). Event-related brain potentials and case information in syntactic ambiguities. *Journal of Cognitive Neuroscience*, *10*, 264-280.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Isel, F., Alter, K., & Friederici, A. D. (2005). Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of Cognitive Neuroscience*, *17*(1), 154–167.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In *Verb-particle explorations* (pp. 67 – 94). Berlin / New York: Walter de Gruyter.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, *284*(1), 1–17.

Kryuchkova, T., Tucker, B., Wurm, L., & Baayen, R. (2012). Danger and usefulness in auditory lexical processing. *Brain and Language*, *122*, 81–91.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*, 223–237.

McIntyre, A. (2007). Particle verbs and argument structure. *Language and Linguistics Compass*, *1*(4), 350–367.

McIntyre, A. (2015). Particle verb formation. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-Formation: An International Handbook of the Languages of Europe* (Vol. 1, p. 435-449). Berlin / Boston: Mouton de Gruyter.

Meinunger, A. (2007). Der Dativ im Deutschen – Eine Verständnishilfe für das Phänomen der gespaltenen Ergativität. *Linguistische Berichte*, *209*, 3–33.

Oldfield, R. C. (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*(1), 97–113.

Olsen, S. (1996). Partikelverben im deutsch-englischen Vergleich [Particle verbs in a comparison of German and English] . In E. Lang & G. Zifonun (Eds.), *Deutsch – typologisch* (pp. 261–288). Berlin: De Gruyter.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428.

Piai, V., Meyer, L., Schreuder, R., & Bastiaansen, M. C. (2013). Sit down and read on: Working memory and long-term memory in particle-verb processing. *Brain and Language*, *127*(2), 296–306.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rohde, D. (2003). Linger: a flexible platform for language processing experiments, version 2.94 [Computer software manual]. http://tedlab.mit.edu/ dr/Linger/.

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, *117*(11), 2341–2356.

Smolka, E., Komlósi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, *24*, 337–375.

Smolka, E., Preller, K. H., & Eulitz, C. (2014). 'Verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language*, *72*, 16–36.

Svenonius, P. (2010). Spatial P in English. In G. Cinque & L. Rizzi (Eds.), *Mapping Spatial PPs. The Cartography of Syntactic Structures* (Vol. 6, pp. 127–160). New York: Oxford University Press.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences. In D. Wood (Ed.), *Perspectives on formulaic language* (p. 151-173). London: The Continuum International Publishing Group.

Urban, S. (2001). *Verbinformationen im Satzverstehen* (dissertation). MPI of Cognitive Neuroscience, Leipzig.

Urban, S. (2002). Parsing verb particle constructions: An approach based on event-related potentials. In *Verb-particle explorations* (pp. 335 – 353). Berlin / New York: Walter de Gruyter.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). *itsadug: Interpreting time series and autocorrelated data using gamms.* (R package version 2.3)

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing chinese relative clauses: Evidence for the subject-relative advantage. , *8*(10).

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth ed.). New York: Springer. Retrieved from `http://www.stats.ox.ac.uk/pub/MASS4` (ISBN 0-387-95457-0)

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83.

Wood, S. N. (2006). *Generalized additive models*. Boca Raton, FL: Chapman and Hall.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3-36.

Woolford, E. (2006). Lexical case, inherent case, and argument structure. *Linguistic Inquiry*, *37*(1), 111-130.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with r.* Berlin, Heidelberg: Springer.

Zwitserlood, P., Bolwiender, A., & Drews, E. (2005). Priming morphologically complex verbs by sentence contexts: Effects of semantic transparency and ambiguity. *Language and Cognitive Processes*, *20*, 395-415.

Notes

[1]For a more detailed theoretical description of particle verbs, and for a distinction between particle verbs and morphologically complex, but nonseparable prefix verbs, see Olsen (1996) and Dehé (2015).

[2]Zwitserlood et al. argue that they found facilitation by words semantically associated with the base verb, but that this semantic facilitation did not work for truly opaque particle verbs without transparent readings. The authors interpret their findings as showing that particle verbs and their bases are associated lexically but not conceptually, in the case of truly opaque particle verbs. Since our own study is concerned with lexical processing load elicited by different verb types in sentence comprehension, we can assume that the distinction between opaque and transparent particle verbs is not relevant for our research question. We will return to this point below in the description of our stimulus material.

[3]Interpreting the results of this study is problematic, mainly because the existing verbs were not matched for frequency, but merely checked for existence in the CELEX database. Equally, ambiguous verbs were not checked to see how plausible their combination with a particle would be. Bearing these concerns in mind, we assume that the general direction of results matches the results of later studies (namely, the N400 at the particle for nonexistent compared to existing base-particle combinations).

[4]We would like to stress that our experiments are not designed to investigate different syntactic proposals for particle verbs, or to decide between different proposals. As outlined above, this would be impossible given the many competing syntactic accounts of particle verbs, and their lack of connection to the sentence processing literature. For an overview on different accounts of syntactic structures and the syntactic status of the objects of particle verbs, we refer the reader to McIntyre (2007, 2015). For an in-depth discussion of the challenges of connecting these syntactic accounts with models of sentence comprehension, we refer to Czypionka and Eulitz (to appear).

[5]Our 56 base verb sample contained four base verbs that only combined to NOM-DAT two-place verbs, und 27 that only combined to NOM-ACC two-place verbs. We will return to these differences in the General Discussion.

[6]The first was *überhören* (to over-hear, 'to ignore something that was said', comparable to 'to overlook'), used in two sentence quartets, which is an existing nonseparable prefix verb, but does not exist as a separable particle verb in German. With the base verb in V2 and the particle at the end, there is no way of interpreting 'hear something over' as 'to overhear something'. The entry given by the dlexDB corpus is associated with this nonseparable prefix verb. The second was *aushören*, 'to out-hear', used in one sentence quartet. A check in the DWDS corpus (the basis for the dlexDB data base) revealed three hits: One from the Kernkorpus in a text from 1919, where the context makes it clear that *aushören* is in fact *aufhören* ('to stop, to end') (a likely typo given the similarity of *s* and *f* in Gothic print), and two in newspaper articles from the 1980ies (Die Zeit); from the context it appears that this was a literal translation from English 'to hear someone out'.

[7]An anonymous reviewer pointed out that many particle verbs are potentially ambiguous when occurring as one complete orthographic unit, and that different readings may come with different subcategorization frames. However, in our stimuli, the particle was the very last word occurring in the sentence, with all potential arguments appearing before. This means that any decision between possible readings would have been made before the particle was presented. We are confident that our pretests allow us to exclude potentially possible alternative readings for naïve participants in our specific stimuli, even if they are possible for the respective verbs in different syntactic structures.

[8]Since our research goals are centered on questions of lexical and syntactic processing, it was not necessary to control the particle verbs used in our stimuli for semantic transparency. As outlined in the Introduction, it has been shown that both semantically transparent and opaque verbs are lexically represented via their bases and particles (Smolka et al., 2009, 2014; Zwitserlood et al., 2005), and are accessed as one lexical unit (Cappelle

et al., 2010). With respect to syntactic differences between conditions, semantic transparency does not interfere with the existence of a verb, or the number of arguments it assigns.

[9]As pointed out by an anonymous reviewer, the predictability of ACC and DAT continuations is especially important for interpreting the results of the planned EEG study. Of the 36 items, 5 had a preference of 74% or higher for DAT continuations, 15 had a preference of 74% or higher for ACC continuations, and 17 items had more evenly distributed preferences. This shows that neither of the two continuations is associated with a much higher predictability, and that the mean percentage for DAT continuations was not driven by a small subset of outlier items.

[10]The difference between the two grammatical conditions accusative and dative on the one hand and the ungrammatical conditions intransitive and illegal on the other hand are facts established by the data (s. Figure 1) and would not need a statistical test for themselves. We cite them here for completeness and readability.

[11]The stop criterion was defined as follows. We estimated the standard deviation $\hat{\sigma}$ of the residuals. From $\hat{\sigma}$ we estimated the absolute value above which the probability for at least one data point falls below 1/2. If we found such data points nevertheless, we removed the largest residual and refit.

[12]We decided to ask for grammaticality judgements instead of comprehension questions because the nature of the stimuli made it extremely difficult to come up with easily answerable comprehension questions - the stimuli included reported speech, ungrammatical sentences etc.. An alternative solution in the literature has been to use a probe detection task (Hopf et al., 2003; Urban, 2001, 2002). We decided against this, since Urban (2001, 2002) reports subtle differences in the outcomes of EEG experiments depending on the behavioural task employed. Importantly, the task employed in these earlier studies did not affect the N400 reported for illegal compared to legal base-particle combinations. Choosing the grammaticality judgment task allowed us some measure of control that participants really performed in-depth parsing of the sentences, and also allowed us to keep the task parallel between the self-paced reading time study and the EEG study.

[13]A very accessible description of more depth can be found in (Baayen, Vasishth, et al., 2016).

[14]Table 1 in 2 gives the actual coordinates.

[15]Following the comments of an anonymous reviewer, we reran our analysis with mean voltages in the 300-500 ms time window. The general pattern of results in this time window was the same as for the shorter time window reported in the main text: There was a statistically significant difference between accusative and illegal conditions ($F = 2.76$, $p < .001$), and a statistically significant difference between accusative and intransitive conditions ($F = 2.25$, $p < .01$). There was no statistically significant difference between accusative and dative conditions ($F = 1.06$, $p = .037$), and no statistically significant difference between intransitive and illegal conditions ($F = .24$, $p = .90$).

[16]The only report of late frontal negativities in the comparable literature is found in Isel et al. (2005). The authors report frontal negativities in the time window from 500-800 ms and 700-1000 ms for conditions containing morpholexical violations in contrast to correct baseline conditions, while no P600 was found for the same contrast. Isel et al. propose that their findings may reflect a "late checking procedure in order to check for a possible combination of the particle and the verb" (Isel et al., 2005, p.162). The late negativities reported by Isel et al. appear to exhibit earlier onset latencies and slightly different topographies than our own (see Isel et al. 2005, p159). We therefore refrain from a decisive interpretation of these frontal negativities.

[17]In general, negative results like the lack of a statistically significant difference between accusative and dative conditions should always be interpreted with caution. However, our experiment design allows us to contrast this negative finding with positive findings from the same participants elicited in the same experimental session. Furthermore, our use of different experimental methods allows us a certain amount of replication. We therefore offer a discussion of our current negative findings for the accusative-dative comparison. An additional goal of this discussion is to inform future experiments aimed to replicate, disprove or extend this finding, to allow for a satisfactory inclusion of lexical case marking in models of sentence comprehension.