The Acceptability of German Noun-Verb Compounds in V2

An Experimental Approach

by

Kyra Kashigin
s1106208

MA degree programme in
Linguistics and Communication Sciences (research)
Department of Language and Communication, Radboud University Nijmegen

Nijmegen, August 2nd 2025
Supervisors: Professor Helen de Hoop
Michelle Suijkerbuijk

Assessor: Dr. Peter de Swart

Radboud University

Acknowledgements

I would like to thank Helen and Michelle for their support, patience and helpful comments on earlier versions of this thesis.
Thank you also to Peter for being second reader and for fun cooperation on other projects during my Master's.
Additional thanks to everyone from *Europe* and *Not Europe* for joking and not joking.
Danke auch an meine Mama für die Unterstützung und die erste Meinung zu all meinen Testsätzen.

Content

Summary

Complex German verbs fall into two categories: prefix verbs which are always inseparable and particle verbs which are split when they occur in verb-second (V2) position. However, there is a group of verbs which has been described as resisting this classification: some complex verbs of the structure N + V – consisting of a base verb prefixed by a nominal element – have been described as being neither separable nor inseparable and instead have therefore been termed as *non-V2 verbs* which are unable to appear in V2 altogether.

Competing theoretical accounts include the attribution of this behaviour to morphological structure (Freywald & Simon, 2007), degree of lexicalisation (Fortmann, 2007) or usage-based factors such as frequency (Forche, 2020). However, empirical support for all of those theories remains insufficient. Previous studies often tested only a small number of verbs with no use of control conditions or full statistical analyses. Disagreement between authors about which verbs should be classified as *non-V2* in the first place further highlights the need for more data and suggests that there may be a high degree of individual variation in speakers' judgements of verbs' ability to appear in V2.

To address these gaps, this study presents a large-scale acceptability judgment experiment that compares 30 potential *non-V2* verbs which have been discussed in the theoretical literature with canonical particle and prefix verbs. Sentences were presented as audio recordings to avoid orthographic bias and participants rated both split and non-split versions in V2 contexts. Linear mixed models and cluster analyses were used to investigate variation at both item and participant levels.

Findings reveal that the tested potential *non-V2* verbs do not constitute a homogenous class. Ratings show significant item- and speaker-level variation with many verbs receiving intermediate scores and some being judged similarly to canonical verbs. The results challenge a classification of the tested verbs as a distinct grammatical category and suggest that individual grammars and a complex interplay of item-level factors better capture the observed patterns. These findings have implications for the classification of complex German verbs, experimental methodology and the treatment of complex verbs in orthographic conventions.

1 Introduction

In main clauses in object-verb (OV) Germanic languages such as German and Dutch, the verb appears in second position as in (1a) so that exactly one constituent fills the position immediately preceding the verb. In contrast, in subclauses the verb stands in clause final position as in (1b).

(1) a. *Der Student schreibt eine Notiz.*
'The student writes a note.'

b. *Der Professor glaubt, dass der Student eine Notiz schreibt.*
'The professor believes that the student writes a note.'

An interesting case is represented by morphologically complex verbs consisting of a verbal stem and a preverbal affix. An example is the verb *zuhören* which consists of the base *hören* 'hear' and the affix *zu* which together form a new verb meaning 'listen'. These verbs are typically analysed as belonging to one of two groups: based on the status of their first morpheme, they are either particle or prefix verbs (Vikner 2005). In the first case, prefix and verb are never separable. In contrast, particle verbs appear either split or non-split depending on the context: in non-V2, the particle appears left adjacent to the verb as in (2a). However, in V2 contexts such as (2b) particle and verb are split with the particle remaining in situ and the base verb moving to V2.

(2) a. *Ich weiß, dass der Chef eine Lösung vorschlug.*
'I know that the boss proposed a solution. '

b. *Der Chef schlug eine Lösung vor.*
'The boss proposed a solution.'

Because of this property, particle verb constructions have received a lot of attention in the theoretical literature where it has been debated whether they should be analysed at the level of morphologically complex words or syntactic phrases (McIntyre, 2015). Most of the particles these analyses have focussed on fall into the class of preposition-like elements or have been described as being homophonous with prepositions (McIntyre, 2002). However, it is also clear that the position preceding the base verb can be filled by more than one preposition as well as other word classes including other verbs, adjectives and nouns (Forche, 2020). It has been argued that these components often behave like particles in that they are separable from their base from which it follows that the whole verb should be analysed as a particle verb (Müller, 2002). However, for a small group of verbs these properties are less clear: those verbs are complex – consisting of a verbal stem and a second morpheme – but appear to be neither separable nor inseparable, i.e. they neither behave like prefix- nor like particle-verbs in that they cannot appear in V2 neither as a whole (3a) nor in part (3b). However, they do appear to be genuine verbs and have been argued to be acceptable in finite non-V2 contexts such as the subordinate clause in (3c) (Fortmann, 2007).

(3) a. * *Er bergsteigt im Sommer.*
he mountain.climb.3SG in summer
'He mountaineers in summer.'

b. * *Er steigt im Sommer berg.*
he climb.3SG in summer mountain
'He mountaineers in summer.'

c. ? *Interessant, dass er im Sommer bergsteigt.*
interesting that he in summer mountain.climb.3SG
'Interesting that he mountaineers in summer.'

Due to this peculiar behaviour, verbs of this kind have been commonly called *non-V2* or *immobile verbs* in the literature (Forche 2020). Many theoretical works have been concerned with the task of identifying such verbs and developing accounts for why they are not mobile. In particular, many different accounts have been postulated seeking the answer in different domains including the verbs' morphology, the semantic relationship between the particle and its base, syntactic properties of German as well as extra-grammatical factors. A major issue of all these accounts is that there is no agreement about which verbs are immobile in the first place. Especially given the low frequency of many of the verbs in question, it remains unclear to what extent the verbs discussed are theoretically relevant at all. The present thesis aims to fill an empirical gap here by examining the acceptability in V2 of the verbs in question. To this end, verbs were selected in a three- step process by consulting both the literature, a corpus as well as native German speakers in a pretest. Next, potential non-V2 verbs were contrasted with canonical separable and inseparable verbs in an acceptability experiment. Results are discussed with reference to their meaning for theoretical accounts of German complex verbs.

2 Particle Verbs in German

In order to understand the problem so-called *non-V2 verbs* pose, it is worthwhile to first provide a more detailed sketch of how canonical complex verbs behave in German. Their most important properties are presented in the following.

Heinz Erhardt, a German post-war stand-up comedian famous for his word plays, once addressed his audience as in (4) (Näser, 1999).

(4) *Sie haben's gut, Sie können sich von Ihren Sorgen **absetzen**, und wir hier oben*
you have it good You can REFL from your sorrows PART.sit und we here up
*müssen uns **einsetzen** damit wir uns **durchsetzen** und Sie nicht **entsetzen***
must REFL PART.sit so that we REFL PART.sit and you not PRE.sit
'It's good for you, you can let go of your worries and we up here have to do our bit to make sure we get our way and don't appal you.'

With this sentence, he demonstrates an important property of German verbs: they may be morphologically complex by consisting of a prefix or a particle followed by a verbal base. Here, he uses four different verbs (in bold in (4)) which all include the root *setzen* 'sit'.

Another time, Erhardt included example (5) in his programme (Näser, 1999) with which he elicited audience laughter following the first sentence and thereby highlighted a second property of verbs in German: they are either separable or inseparable.

(5) *Es lüstet mich ge… Es gelüstet mich.*
it ROOT me PREF It PREF.ROOT me
'I have a craving.'

There is nothing about the semantic content of (5) that is funny in this context. Instead, the humour is entirely derived from the fact that *gelüsten* 'have a craving for something' is a non-separable complex verb used in its split form which is unacceptable to speakers. The audience's laughter thus proves to be a live acceptability judgement showing that the utterance is unacceptable to speakers. In syntactic notation this utterance should thus be written as (6).

(6) a. * *Es lüstet mich ge.*
b. *Es gelüstet mich.*
'I have a craving.'

The first category of complex German verbs which *gelüsten* is part of are so-called *prefix verbs*. In this category, the prefix stems from a limited group of the six morphemes: *ent-*, *er-*, *ge-*, *ver-*, *be-* and *zer-* (Duden, 2009). These prefixes never occur as lexical items of their own, are unstressed and can never be split from their base as in (7) (ibid.). In syntactic analyses they can thus straightforwardly be treated as being akin to verbs with a simplex stem in that all syntactic and morphological operations prefix verbs undergo operate on the level of the entire word.

(7) a. * Mein Nachbar kauft sein Auto ver.
my neighbour buy his car PREF
b. Mein Nachbar verkauft sein Auto
my neighbour PREF.buy his car
'My neighbour sells his car.'

Separable complex verbs, on the other hand, are prefixed by a particle of various degrees of semantic transparency. These are typically derived from the class of prepositions (McIntyre, 2002) and examples include *durch* 'through', *nach* 'after' or *ab* 'of'. The particle always bears the word accent (Biskup, Putnam & Smith 2011) and the number of possible particles for separable complex verbs is much larger than that of inseparable prefixes which is why no exhaustive list can be given here. Although the terms are occasionally used interchangeably, the standard convention in the literature is to call separable verbs *particle verbs* and non-separable verbs *prefix verbs* which is also the terminology adopted throughout this thesis.

In contrast to prefix verbs, particle verbs are separable: as previously shown in (2), particle verbs appear syntactically split in V2 but not in non-V2 contexts. Furthermore, both the past participle affix *ge-* (8) and the infinitival marker affix *zu* (9) intervene between particle and its base. This affixation is sometimes called *morphological separability* (cf. e.g. Forche, 2020) in order to distinguish it from the separation in V2 which is termed *syntactic separability*.

In addition to this, it has also been argued that it is possible to topicalise the particle as in (10). The acceptability of this fronting has been experimentally shown to depend on the semantic transparency in the relationship between base and particle as well as on whether the particle is used contrastively (Trotzke, Quaglia & Wittenberg, 2015).

(8) *Toms Mutter hat eine Geschichte vorgelesen.*
Tom.POSS mother has a history PART.PTCP.read
'Tom's mother has read out a story.'

(9) *Gib mir Zeit nachzudenken.*
give me time PART.INF.think
'Give me time to think.'

(10) *Auf gibt er nie.*
up give.3SG he never
'He never gives up.'

The obvious theoretical problem posed by particle verbs is that their properties lie at the interface between syntax and morphology which makes it not straightforward to assess whether the particle verb construction represents a single word or two. The facts that only the base verb moves to V2 and that the particle may be fronted without its base are taken as the strongest evidence for a syntactic analysis. More specifically, under such a view, base and particle are assumed to be two separate words with the particle projecting a syntactic phrase on its own and possible a direct object if there is one (Dehé, 2015). In contrast, evidence supporting a complex word analysis is the fact that the particle verb construction as a whole can undergo word formation processes as in (11) (McIntyre, 2007) as well as particle verbs' often untransparent meaning (Müller, 2002) which in turn is interpreted as evidence for particle and verb being stored as one unit in the mental lexicon (ibid.).

(11) ANziehsachen 'on.put.things = clothes'

Evidence for the status of these verbs also comes from processing studies. For instance, Czypionka et al. (2019) show that encountering a particle in sentence-final position triggers both syntactic and morphological integration, as represented by an enhanced N400 reflecting lexical access as well as a P600 indicating syntactic integration efforts.

While the majority of particle verbs consist of a verb preceded by a prepositional element, the first position in a complex verb is by no means limited to a single word from this category. Instead, this position may be filled by more than one preverbal element (12a), adjectives (12b), other verbs (12c) as well as nouns (12d). As will be seen in the following sections, some of these constructions create challenges for the binary grouping of complex German verbs into either prefix or particle verbs.

(12) a. *Die Gäste wollen sich für die nächste Veranstaltung voranmelden.*
the guests want REFL for the next event PART.PART.register
'The guests want to preregister for the next event.'

b. *Der Arzt schreibt seinen Patienten krank.*
the doctor write his patient sick
'The doctor writes a sick note for his patient.'

c. *Der Bauer muss zum Ernten mähdreschen.*
the farmer must to harvest mow.thresh
'The farmer has to thresh to harvest.'

d. *Die Lehrerin muss die Klausur korrekturlesen.*
the teacher must the exam correction.read
'The teacher has to correct the exam.'

3 *Non-V2* verbs

While most complex German verbs can be grouped into either *separable* or *inseparable* verbs, a small number of verbs has received considerable attention in the literature for appearing to fall into neither of those categories. So-called *non-V2 verbs* are always complex and have been argued to be neither separable nor inseparable. As already demonstrated in (3a-c), this means that these verbs neither appear in V2 as a whole (as prefix verbs do) (3a) nor split from their particle (as particle verbs do) (3b) making it impossible for them to appear in V2 altogether. In contrast, it is generally assumed that their appearance in non-V2 contexts is perfectly acceptable as in (3c) (Fortmann, 2007) making them full verbs nonetheless (Ahlers 2021). The most commonly discussed verbs that are said to fall into this category are either of the internal structure Prt + Prt + V such as *voranmelden* 'preregister' in (13a) or N + V such as *korrekturlesen* 'correction.read' = 'proofread' in (13d). While the former category is generally agreed upon to be non-V2, there is much more variation in the assessment of how N + V verbs behave syntactically. That means that while most non-V2 verbs are N + V compounds, not all verbs falling into this category exhibit the same behaviour (Fortmann, 2007).

The present research will thus be concerned with disentangling the variation in verbs consisting of a verbal base preceded by a nominal element. Given the amount of often contradictory analyses in the literature, it will then be argued that speaker data is the key to developing a thorough understanding of these verbs.

3.1 Theoretical Accounts

A range of different accounts have been proposed to explain potential non-V2 N + V verbs' behaviour. The overview provided here is by no means exhaustive but rather intends to illustrate the complexity of the problem by showing how it pertains to morphology, syntax, semantics and their interfaces.

3.1.1 Morphology

Morphological accounts deal with an important property of German verbal compounds consisting of a noun and a verb: they are not supposed to exist. As outlined by Ahlers (2021), it has been postulated since Grimm (1878) that verbal compounds of the structure N + V are not productive in German. As seen previously, however, exceptions do exist which is why special mechanisms for how those verbs emerged have to be postulated. Two options which both assume denominalisation from a noun to a verb are *backformation* and *conversion*. In the process of backformation (Fortmann, 2007) a verb can emerge from a noun via deletion and replacement of a nominal suffix such as in (13) while in the case of conversion the noun is verbalised directly without a morphological change of the surface form such as in (14). The results of both of these types of word formation processes have also been described as *pseudo-compounds* (Freywald & Simon, 2007) which illustrates that these items are not assumed to have emerged from compounding a noun and a verb but rather from turning an existing nominal compound noun into a verb.

(13) *Notland~~ung~~* (N) → *Notlanden* (V)
emergency landing (to) emergency land

(14) *Bausparen* (N) → *Bausparen* (V)
building saving (to) building save

When it comes to the syntactic behaviour of complex verbs whose first component belongs to the category of nouns, it has been proposed that their internal morphology and particularly their derivational process is responsible for their V2 restriction: because these verbs are not actually verbal compounds, they cannot stand in V2. Problems with this account are that it remains unclear why exactly speakers would be restricted by the internal diachronic word origin and the fact that this claim contrasts with the *Lexical Integrity Hypothesis* (Di Sciullo & Williams, 1987) which states that a word's morphology cannot have an effect on its syntactic behaviour (Freywald & Simon, 2007).

A third verb formation process which is productive in many languages is that of *noun incorporation* where a noun becomes incorporated into the verbal stem to form a new verb. In this process, the noun typically loses its referentiality and individual modifiability which is why the resulting verb is often used to denote habitual actions (Gerdts, 2017). In most Germanic languages, however, incorporation is assumed to be unproductive. Booij (2008) argues that cases that appear as incorporation must therefore be described as *pseudo-incorporation*. For Dutch he describes constructions where a nominal component can be read as an argument of the verb, such as *piano spelen* 'play the piano'. If such a reading is available, the construction can stand in V2 with the nominal component split from the verb as in (15) taken from Booij (2008: 5).

(15) *Jan speelt piano.* (Dutch)
Jan plays piano
'Jan plays the piano.'

One way in which this type of pseudo-incorporation can be distinguished from regular verb plus object combinations is the progressive construction. For example, in Booij's (2008: 6) examples of *piano spelen*, both (16a) and (16b) are acceptable while for regular objects, only the position preceding the progressive marker *aan het* 'at the' is possible. This highlights the close union between noun and verb in the pseudo-incorporated construction.

(16) a. *Jan is aan het piano spelen.* (Dutch)
Jan is at the piano play.INF

b. *Jan is piano aan het spelen.*
Jan is piano at the play.INF
'Jan is playing the piano.'

The same pattern seems to largely hold true for German as well, though exceptions exist. Specifically, cases like *manndecken* 'man.cover' and *ehebrechen* 'marriage.break' = 'commit adultery' are widely reported to be non-V2 (cf. e.g. Fortmann, 2007) even though they appear to be pseudo-incorporated verbs where the nominal and the verbal component stand in a verb-argument relationship which means they should be separable according to Booij's theory.

Booij (2008) further predicts that in cases where the nominal component does not function as an argument, the verbal complex must be reanalysed as a stem in order to be able to appear in V2. For example, in German, *ohrfeigen* 'ear.sweep' = 'slap someone in the face' is frequently cited as unproblematic in V2 contexts (e.g. Fortmann, 2007; Ahlers, 2020) which is likely because the verb has undergone reanalysis as a single lexical unit with the stem *ohrfeig-*. This reanalysis allows the verb to maintain its syntactic integrity without requiring separation between noun and verb. If such a reanalysis does not occur, however, the verb cannot occupy V2, neither as a whole nor in its split form.

In sum, Booij (2008: 9) assumes three types of N + V constructions as in (17). The first option represents a regular verb-argument relationship, where the nominal component is a DP that can be independently modified such as in the German example in (18), where the object *der Bus* 'the bus' is modified by the adjective *früh* 'early'. In the other two cases, N and V form a unit either at V – making them inseparable, such as *ohrfeigen* – or at V' which guarantees their separability as in the case of the construction *Bus fahren* 'Bus drive' = 'take the bus' where the noun is no longer independently modifiable as in (19). If none of these readings is available, the verb will not be able to appear in V2.

(17) regular syntax: $[[N]_{DP}\ V]_{VP}$
pseudo-incorporation: $[N\ V]_{V'}$
compounding: $[N\ V]_{V}$

(18) *Sie nimmt den frühen Bus zur Arbeit*
she takes the early bus to work
'She takes the early bus to work.'

(19) *Sie fährt immer {Bus / * frühen Bus} zur Arbeit*
she drives always Bus early bus to work
'She always takes the early bus to work.'

In addition to these options, Eschenlohr (1999) argues that at least in German, there is another possible outcome in which the noun can be split from the verb: if an object reading is unavailable, the nominal component may be reanalysed as a particle, which would then allow the verb to appear in V2 position like a regular split particle verb. For example, Eschenlohr (1999) claims that *kopf* 'head' functions as a particle in *kopfstehen* 'head.stand' but not in *kopfrechnen* 'head.calculate'. She further argues that such a reanalysis is more likely if the noun is compatible with a range of verbs which she describes as nouns that are *reihenbildend* 'row building'. However, this is a theory that still does not sufficiently explain in what cases the entire compound would be reanalysed as a stem and when only the nominal element would be reanalysed as a particle.

3.1.2 Semantics

An approach based on the semantic relationship between noun and verbal base is brought forward by Fortmann (2007), who includes an analysis of a range of possible semantic relationships between the nominal and the verbal component. He broadly differentiates two classes of N + V compounds: in the first one, the noun and verb stand in a transparent relationship to each other. The noun may either be an argument of the verb, as in Booij (2008) or its modifier as in the verb *schutzimpfen* 'protection.vaccinate'. In this case, as Fortmann argues, *Schutz* 'protection' functions as a type of final modification with the meaning of 'to vaccinate with the purpose of protection'. Verbs of this group are therefore semantically transparent and compositional.

In contrast, a second group of verbs includes those for which the relationship between N and V is non-compositional and untransparent. Examples include *maßregeln* 'measure.regulate' = 'discipline' where the noun neither functions as an argument nor modifies the verb in any discernible way. These verbs are idiomatic and their meaning cannot be derived from their individual components. Importantly, Fortmann argues that this distinction has syntactic implications: due to their untransparency, idiomatic verbs can be reanalysed as unitary stems which can then appear in V2 as a whole just like prefix verbs.

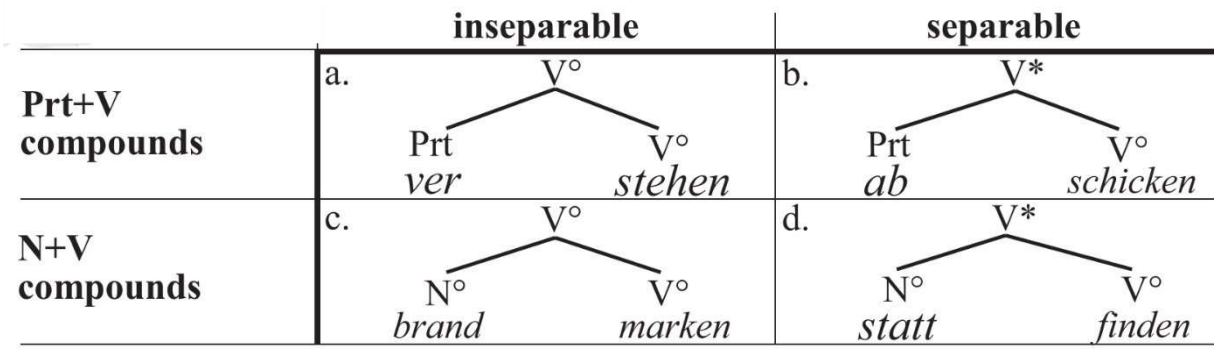
Fortmann also identifies a possible connection between verbs' internal morphology and their semantics: he argues that in verbs originating from backformation such as *schutzimpfen* 'protection.vaccinate' via *Schutzimpfung '*protection.vaccination' the noun retains semantic control over the verb, which blocks syntactic movement in the case of transparent verbs. Conversely, verbs like *maßregeln* 'measure.regulate' = 'discipline' which do not display this dependence are not subject to such constraints.

3.1.3 Syntax

In a syntactic account, Vikner (2005) postulates different projections for the different types of complex verbs: he analyses prefix verbs as V° corresponding to the same projection as that of simplex verbs. In contrast, separable particle verbs in Standard German as well as Dutch and Swiss German are classified as V* – a projection between V' and V° – ensuring the separability of these items. These structures are depicted in Figure 1 taken from Vikner (2005: 93). As visible here, Vikner shows that both separable and inseparable N + V compounds exist.

Figure 1

Types of German complex verbs able to stand in V2 according to Vikner (2005)



In addition to regular particle and prefix verbs, Vikner postulates verbs which cannot appear in V2 as a third category of complex verbs. In his theory, the problem of non-V2 verbs then exists because these are verbs which have to fulfil both the requirements of verbs at V° and verbs at V* at once. That is, they are subject to the pressure to be split (like V* verbs) while also being required to move to V2 as a whole (like V° verbs). This dual requirement creates a structural conflict that renders the verbs ungrammatical in V2 regardless of which structure is chosen.

Vikner however remains very unclear about why this might be the case. His only suggestion is that verbs may be required to be split due to consisting of two clearly identifiable parts but that the requirement to stand in V2 as a whole is also imposed on them because they are often turned into a verb as whole by verbalising a noun. This results in an insolvable structural conflict. This account, however, remains very vague and does not make predictions about which N + V verbs are affected by this conflict.

An aspect that is clearer about Vikner's theory, is that it also accounts for variation between OV/VO Germanic languages: in contrast to OV languages such as German and Dutch, there cannot exist a structural conflict in V2 in VO languages because of the order in which particle and verb are aligned. In Danish, a Germanic VO language, for example, inseparable verbs at V° always follow the pattern

particle-verb which is the same as in German. In contrast, this order is reversed to that of verb-particle for separable verbs at V* which mirrors the VO word order. This can be seen by comparing (20a) taken from Vikner (2005: 109) with its German translation in (20b). Hence, any novel verb in Danish only has the option of following one of those structures and no conflict between them can arise.

(20) a. *Hvorfor kunne konferencen ikke finde sted her?* (Danish)
why can conference.the not take place here

b. *Warum kann die Konferenz hier nicht stattfinden?* (German)
why can the conference here not place.take
'Why can't the conference take place here?'

Unlike other authors such as Fortmann (2007), Vikner also acknowledges that his account may not always make clearcut predictions about the V2 status of any given verb. Instead, he argues that there may be substantial individual variation between how speakers treat the items in question. In formal terms, this means that for some speakers potential immobile verbs may have developed either a V° or V* projection and are therefore acceptable to them in V2.

3.1.4 Usage-based Accounts

In line with Vikner's suggestion, extragrammatical performance factors which include novelty and frequency of the lexeme have also been proposed as explanatory variables by other authors. In particular, the less frequent and more novel a verb is, the less likely it is supposed to have developed into either a prefix or a particle verb and the more uncertainty speakers are supposed to show in using it (Freywald & Simon, 2007). This is theoretically supported by the postulation that new verbs develop as schematised in (21) adapted from Stiebels & Wunderlich (1995: 946): under the assumption that the last stage in full verb development is the verb's finite form in V2 contexts, verbs may already be usable in other contexts while still causing uncertainties in V2.

(21) infinitive → zu-infinitive → finite verb: clause-final → finite verb: V2

Similarly, frequency may also affect the usage of lexemes when the word is old. For instance, Forche (2020) argues that the previously discussed *ehebrechen* 'marriage.break' = 'commit adultery' is mostly used in biblical and otherwise religious contexts and is outdated nowadays. As a consequence, the verb occurs much less frequently leading to it not being used in its full inflectional paradigm anymore. In usage-based accounts, it is then not so much the internal composition of the verb itself which causes its peculiar behaviour but rather the rareness of these types of items which causes uncertainty about their syntactic behaviour.

What makes weighing between all of the discussed proposals more difficult is that judgements about which verbs are non-V2 differ between authors. What is arguably the most extreme case is uncovered by Freywald & Simon (2007) and found in Eisenberg (2004) who reports *weil sie bauspart* 'because she building saves' as ungrammatical but *weil er bauspart* 'because he building saves' as grammatical within the same work. Notable cases like this demonstrate that it is unclear how much data any of the theories reviewed above should account for because not all constructions theoretically treated as problematic might indeed be problematic to speakers. It is therefore clear that introspective judgements from individual researchers do not suffice. Verbs like these must instead be approached with empirical data.

3.2 Empirical Work

Despite the long-standing interest in non-V2 verbs in the theoretical literature and the resulting clear need for empirical data, well-developed studies investigating the status of potential non-V2 verbs are rare and many of them are more exploratory in nature and include neither a fully controlled experimental design nor a reliable statistical analysis. However, a few studies of German are relevant here and can serve as the basis on which to build a more controlled study. An overview of empirical results for N + V compounds is provided in Table 1. Each work will be discussed in more detail in the following.

Table 1

List of all items which have been examined experimentally and their attested acceptability. Verbs in bold have previously been argued in the literature to be non-V2 (cf. Forche, 2020).

Verb	Study	Result split	Result non-split
bausparen	Freywald & Simon (2007)	1.2 (0 – 6 scale)	4.3 (0 – 6 scale)
ehebrechen	Freywald & Simon (2007)	1.2 (0 – 6 scale)	1.1 (0 – 6 scale)
notlanden	Frewald & Simon (2007)	0.8 (0 – 6 scale)	1.7 (0 – 6 scale)
schutzimpfen	Forche (2020)	1.8 (0 – 6 scale)	0.9 (0 – 6 scale)
bausparen	Forche (2020)	0.3 (0 – 6 scale)	2.6 (0 – 6 scale)
raubkopieren	Forche (2020)	0.5 (0 – 6 scale)	4.7 (0 – 6 scale)
stoßlüften	Forche (2020)	2.2 (0 – 6 scale)	4.1 (0 – 6 scale)
notoperieren	Forche (2020)	2.3 (0 – 6 scale)	4 (0 – 6 scale)
arschkriechen	Ahlers (2021)	*NA*	3.3 (0 – 4 scale)
ratschlagen	Ahlers (2021)	*NA*	2.9 (0 – 4 scale)
bergsteigen	Ahlers (2021)	*NA*	3.6 (0 – 4 scale)
lobpreisen	Ahlers (2021)	*NA*	1.8 (0 – 4 scale)
ohrfeigen	Ahlers (2021)	*NA*	3.5 (0 – 4 scale)
handhaben	Ahlers (2021)	*NA*	3.5 (0 – 4 scale)
ehebrechen	Terboven (2021)	28.6% acceptable	26.2% acceptable
bergwandern	Terboven (2021)	0% acceptable	69% acceptable
maßregeln	Terboven (2021)	19% acceptable	95.2% acceptable
eislaufen	Terboven (2021)	57.1% acceptable	47.6% acceptable
wertschätzen	Terboven (2021)	58.5% acceptable	85.7% acceptable
notlanden	Terboven (2021)	90.5% acceptable	47.6% acceptable
ohrfeigen	Terboven (2021)	0% acceptable	100% acceptable
schlangestehen	Terboven (2021)	100% acceptable	0% acceptable
Klavier spielen	Terboven (2021)	100% acceptable	0% acceptable

3.2.1 Freywald & Simon (2007)

In an initial exploration, Freywald and Simon (2007) examine three N + V verbs frequently discussed in the literature: *bausparen* 'building.save', *ehebrechen* 'marriage.break = 'commit adultery', and *notlanden '*emergency.land'. These three items were chosen because they are argued to represent the three ways morphologically complex N + V verbs can emerge in German: conversion (*bausparen*), (pseudo) incorporation (*ehebrechen*), and back-formation (*notlanden*).

In an acceptability judgment task, participants were asked to rate the unusualness of these items in finite clause-final position, as well as in their split and non-split forms in V2 position. In addition to the three N + V items, the authors included two Prt + V items, which are thought to have emerged through

the same word-formation processes: zwischenspeichern ‘between.save’ (conversion) and zwischenlanden ‘between.land’ (back-formation).

Due to the small number of items, only descriptive results are presented, making it difficult to draw reliable conclusions. However, since verbs with a preposition as their first component are much more acceptable in V2 position than their N + V counterparts, the authors’ main interpretation is that the much-discussed word-formation processes are not responsible for the V2 status of verbs. Instead, it appears that the category of the first component is decisive in determining whether a verb might be non-V2 or not, with N + V verbs being particularly prone to unusual behaviour in this regard. These results underline the relevance of further research on N + V verbs.

3.2.2 Forche (2020)

As the main body of his dissertation, Forche (2020) conducts an extensive corpus study of potential non-V2 verbs. He presents records of the syntactic contexts in which his target verbs appear and contrasts those with the position of canonical verbs.

His most important finding is that the target verbs only appear in their finite form in about 15% of the data compared to canonical verbs for which this is the true in more than 50% of cases. From this, he draws the conclusion that potential non-V2 verbs are largely non-finite. According to him, this observation would explain the perceived non-V2-ness of verbs: in this theory, verbs are not specifically non-V2 but are rather unable to appear in their finite form at all which would then include but not be specific to V2 contexts. This argumentation converges with Freywald & Simon’s (2007) account according to which verbal paradigms develop over time and verbs first emerge in their infinitival form. This means the examined verbs would simply not have developed a finite inflectional paradigm yet, following Stiebels and Wunderlich’s (1995) proposed order of verb development (cf. (22)). Forche statistically corroborates his finding in a logistic regression model which shows that bearing a double particle is the only variable exclusively predicting that a verb becomes non-V2. In contrast, all other significant predictors of an increased chance of non-V2-ness also significantly predict a verb’s non-finiteness.

When it comes to what these predictors are, Forche shows that the morphological factor of backformation is not a significant predictor. Instead, verbs’ appearance in V2 appears to be mostly restricted by semantic properties. First of all, verbs which express a telic action are less likely to appear in V2. The same applies to verbs which, as Forche argues, are oriented towards the Patient such as *schutzimpfen*. Forche explains this by assuming that this category of verbs typically appears in passive constructions. This would thus mean that the verb is not directly blocked from appearing in V2 but is simply less likely to do so due to its semantics.

Importantly, transparency in the relationship between noun and verb significantly *increases* a verb’s chance of appearing in V2 in Forche’s model. This strongly contrasts with Fortmann’s (2007) theory which predicts the exact opposite behaviour. However, as Forche concedes, it must be noted that his analysis only included verbs which have been assumed to be non-V2 in the literature. This means that verbs cited by Fortmann as examples for N + V compounds which can be reanalysed as unitary stems due to their untransparent non-compositional nature are not included in the dataset which may be why this effect does not crystalise in Forche’s data.

Results from the corpus study are complemented by Forche by a series of behavioural experiments. In an acceptability judgement task, he asked participants to rate the acceptability of items in V2 on a five-point scale. In another task, participants were instructed to insert target verbs with the option being open of the verb being inserted in its split or non-split form. Results of both experiments show a clear tendency only for complex verbs possessing more than one particle but not for N + V compounds. While – in line with his corpus study – the first category appears to be strongly non-V2, the picture is much less clear for N + V compounds which receive high ratings and rate of occurrence in V2 for verbs such as *raubkopieren* 'theft.copy' = 'pirate' and very low ratings for other such as *schutzimpfen*. Forche (2020) does, in fact, also argue that there are verbs which had previously been argued to be non-V2 but which do in fact show finite occurrence in V2 in the corpus data. This suggests that some of the verbs theoretically described as being problematic are in fact fully functioning verbs. Just like Freywald & Simon's work, these findings corroborate the need for more empirical data about N + V compounds as they suggest that some of the verbs discussed in the theoretical literature are not non-V2 at all.

3.2.3 Ahlers (2021)

According to Ahlers (2021), verbs' non-V2 property is explained by syntactic processing constraints. In order to show this, he combines an acceptability judgement task with a task asking participants to form the past participle of verbs for six verbs prominent in the literature: *ohrfeigen* 'ear.sweep' = 'slap in the face', *handhaben* 'hand.have' = 'handle', *ratschlagen* 'advice.hit' = 'consult', *arschkriechen* 'ass.crawl' = 'boot lick', *bergsteigen* 'mountain climb' and *lobpreisen* 'laud.praise'.

What differentiates the verbs used by Ahlers, is that while all of them are N + V compounds, only the latter four have been argued to be non-V2 verbs (cf. Forche, 2020) while the other two are typically regarded as being able to appear in V2. Ahlers (2021) explains this difference with what he calls the head inflection hypothesis (*Kopfflexionshypothese*): according to this theory, N + V verbs can stand in V2 if the entire compound is reanalysed as the verbal stem. If, however, this does not happen, the verb will be inflected on the internal head which makes it immobile.

Ahlers argues that speaker data confirms his hypothesis: firstly, in the acceptability judgement task, participants rated the target verbs in V2. Descriptive results as presented in Table 1 show ratings of between 2.9 and 3.6 on a scale from 0 (fully unacceptable) to 4 (fully acceptable) for the verbs *ratschlagen*, *arschkriechen* and *bergsteigen*. This suggests that some speakers do in fact find them to be acceptable in V2. Only *lobpreisen* scored lower than 2.

In the past participle formation task, Ahlers aimed to capture verbs' morphological separability by testing whether speakers use the past participle affix *ge-* as a prefix of the entire compound or only of the verbal stem. Results show that the V2 (mobile) verbs *ratschlagen* 'advice.hit' = 'consult', *handhaben* 'hand.have' = 'handle' and *ohrfeigen* 'ear.sweep' = 'slap in the face' behave as expected of them. That is, more than 90% of speakers prefix them in their past participle form. In contrast, there is considerable variation for the other three verbs in that they appear to be morphologically separable only for some speakers. Using the variable of whether speakers pre- or interfixed a verb as a predictor, Ahlers then links the past participle to the acceptability judgement experiment: specifically, he shows statistically that speakers more readily accept complex verbs in V2 if they reanalyse the entire compound-like structure as the verbal stem and form the past participle via pre- rather than interfixation. These results are in line with what Fortmann (2007) describes in that they seem to confirm that a verb's mobility depends on whether it can be reanalysed as a simplex stem or not. Additionally, they agree with Vikner's (2005) predictions of individual variation because they show that speakers use different strategies in dealing with the tested verbs. Nevertheless, Ahlers does not

provide a more detailed analysis of this variation.
In addition, unlike Fortmann, Ahlers (2021) also expands his theory to account for why complex N + V verbs which are not fully reanalysed as the verbal stem cannot appear in V2 in their split form: extending from Eschenlohr (1999), he suggests that if verbs are inflected on their internal head, any element bearing a word accent preceding it has to be split from it since only one element can precede the verbal stem in V2 environments. Verbs that are split, however, violate processing expectations in that the parser expects a particle from a closed lexical class or an argument of the verb. Either option is unavailable in non-V2 contexts, which is why those verbs are available there in their finite form.

Despite this account, Ahlers' work does not actually provide any experimental data to support his claim that the tested verbs are also unacceptable in V2 contexts when split and since not all of Ahler's used verbs are supposed to be non-V2, results are only presented for four verbs. Hence, it remains unclear to what extent results can be generalised and extended to other verbs.

3.2.4 Terboven (2021)

Another publication which does test target verbs' ability to appear in V2 in their split form is Terboven (2021). In his dissertation, the author selects nine verbs which – as in Freywald & Simon (2009) – are supposed to represent different categories of internal morphology. Out of these nine items, three have previously been argued to be non-V2 verbs (cf. Forche 2020) while the others are standardly assumed to be able to stand in V2 either split or as a whole.

In a binary acceptability judgement task, participants were asked to judge whether the target verbs are acceptable or unacceptable in V2 as a whole and in their split form. As shown in Table 1, results reveal that the potential non-V2 verbs *bergwandern* 'mountain.climb' and *notlanden* 'emergency.land' receive acceptance rates of over 50% in V2. This implies that some speakers are in fact able to use variants of the target verbs in V2 position. In another experiment, participants were asked to decide whether they would add the infinitival marker *zu* as a prefix or interfix to the complex verb. This experiment showed whether the verb is morphologically separable for speakers or not and is thereby similar to Ahlers' participle formation experiment.

Linking these tasks, Terboven groups verbs into three categories: firstly, verbs which are neither morphologically nor syntactically separable are *handhaben* 'hand.have' = *'handle'* and *ohrfeigen* 'ear.sweep' = 'slap in the face'. These verbs behave like canonical non-separable prefix verbs which is in line with theories which assume a reanalysis of semantically intransparent verbs as a simplex stem. Secondly, there are verbs which are morphologically but not syntactically separable. This category includes only hypothesised non-V2 verbs, namely *eislaufen '*ice.walk', *bergwandern '*mountain.hike', *notlanden* 'emergency.land', and *wertschätzen* 'worth.value'. Thirdly, there are some verbs which are both morphologically as well as syntactically separable which include *ehebrechen* 'marriage.break' = 'commit adultery', *schlangestehen* 'line.stand' = 'queue', and *klavierspielen* 'piano.play'. These are all pseudo-incorporated verbs which appears to confirm Booij's (2007) hypothesis.

Terboven (2021), however, only tests very few target verbs, does not use control items, uses a binary rating scale and does not explore individual variation. Such a procedure goes against gradient approaches of acceptability in general and against the idea that speakers display uncertainties in the use of non-V2 verbs in particular. Research shows that a range of factors including sentence aesthetics, prescriptive rules and perceived probability affect sentence judgements (Schoenmakers, 2023) and

that when using a more nuanced rating scale, speakers display variability in their rating behaviour (Schoenmakers & van Hout, 2024). Especially given the fact that there is large disagreement and uncertainty between theoretical authors, this variation and gradience should be attempted to be captured.

4 Current Research Gaps

From the reviewed studies, research gaps can be identified, which are both theoretical and methodological in nature. In total, the studies presented here have examined fifteen different potential non-V2 verbs. This small number contrasts greatly with the number of over 1000 potential non-V2 verbs Forche (2020) identified in a literature review and underlines the necessity for further research. Patterns found in Ahlers (2021) and Terboven (2021) are only attested for a very small number of verbs and can therefore not be generalised. Thus, all of the previous data has to be compared to a larger amount of data.

Besides the necessity for more data, this should also be collected more carefully by conducting a more controlled study. None of the previous explorations used control conditions. This is, however, strongly necessary if one wants to draw conclusions about the extent of the anomaly of the verbs in question because it has to be known to what extent the target word acceptability differs from that of canonical verbs.

A further methodological issue not addressed in any of the previous studies is that of modality: all prior experiments have presented target items to participants in written form. This poses a problem because N + V compounds are a frequent topic in discussions of standard German orthography. Since German orthography capitalises nouns, the decision to treat these compounds as either a single verb or a syntactic unit consisting of a noun and a verb directly affects their spelling. Due to their status at the intersection of morphology and syntax, orthographers have struggled with them—so much so that the officially recognized spelling of the word for ice skating was changed from *eislaufen* to *Eis laufen* and back within just ten years between 1996 and 2006 (Zwischenstaatliche Kommission für deutsche Rechtschreibung, 2004; Rechtschreibrat, 2024). Even syntactic theorists, such as Müller (2002), admit to having been influenced by orthographic conventions in their past analyses which, for instance, led to his classification of *Rad fahren* ‘bike drive‘ = ‘cycle’ as two words.

That German speakers’ sentence processing is, in fact, sensitive to deviations from standard capitalisation is demonstrated by Labusch et al. (2022), who showed that deviations from the standard spelling of first-letter capitalisation in German affect lexical access in speakers. In the case of particle verbs, this means that using upper- or lower-case spelling at the position of the particle-like element might bias speakers towards either a particle or a noun reading, resulting in low acceptability when that reading does not match the rest of the sentence or the status this verb has for that particular speaker. Additionally, data from Ahlers (2021) shows that when asked to freely enter the past participle of certain verbs, some speakers add a space between the particle-like element and the verb, suggesting that they analyse them as two words rather than as one verb. Specifically, speakers’ spelling is affected by their interpretation of the grammatical status of the N + V construction: in another recent publication by Schäfer and Sayatz (2024), the authors examine the orthography of N + V constructions in German and argue that the graphemic variation seen in speakers’ spelling of these constructions can be explained through a usage-based probabilistic grammar model. This model posits that the spelling choice reflects underlying grammatical structures. Specifically, speakers are more likely to spell an N + V construct as two words when the two elements are perceived to be in a predicate/argument

relationship. A final point of concern about presenting experimental items in writing is Forche's (2020) suggestion that N + V compounds may simply be more frequent in spoken and spontaneous language. Because of all of these issues, there is a need for a study presenting items in a more natural context which does not bias participants towards one interpretation or another.

A more theoretical problem with some of the previous experimental setups, is that creating experimental conditions for different types of potential non-V2 verbs does not appear to be sensible. This is because previous empirical findings are not informative enough and all theoretical accounts reviewed here have been challenged with counterexamples. Furthermore, it often seems difficult to fully tell for individual lexical items what any given theory would predict about them. For example, as described by Forche (2020), it is not clearly identifiable whether the verb *ehebrechen* 'marriage.break' = 'commit adultery' is derived via backformation from the noun *Ehebruch* 'marriage.break' = 'adultery', or whether it is a product of pseudo-incorporation of the noun *Ehe* 'marriage'. These two possibilities entail different predictions about the verb's V2 status: while backformation predicts that the verb cannot appear in V2, pseudo-incorporation, as described by Booij (2008), makes the opposite prediction. In addition, many theories would predict clear cases of either prefix or particle verbs but a quick look at the literature suffices to see that there is a lot of variation in how speakers judge these cases. There is thus ample room for a methodologically improved study on German N + V constructions.

5 The Present Study

With the findings and shortcomings from previous research in mind, the goal of the present study is to empirically capture the gradience in acceptability within and between items as well as between speakers for N + V verbs which have been discussed in the literature. The only assumption to be held is that in principle, there are three main pathways to go: those of particle verbs, of prefix verbs and of non-V2 verbs. The question of which path individual verbs take is subject of this study. To this end, the following research questions are asked:

(RQ1) How do native speakers of German rate the acceptability of potential non-V2 N + V verbs in V2 position compared to canonical particle and prefix verbs in the same contexts?

(RQ2) What patterns can be observed in the acceptability ratings of target verbs, and how do these patterns relate to factors such as morphological structure, semantic properties, and frequency?

(RQ3) What patterns emerge in how individual speakers rate the different target verbs?

The first research question is concerned with identifying non-V2 verbs as perceived by speakers. This question aims to provide a comprehensive basis for the following analysis which is more reliable than judgements of individual researchers. It is hypothesised that particle and prefix verbs will behave as expected, that is, prefix verbs should be highly unacceptable in their split form in V2 contexts and highly acceptable in their non-split form while the exact opposite pattern should be observed for particle verbs. Results for the target verbs, ill then show how these items pattern in relation to the other two categories. Given the disagreement in the literature, considerable variation between items as well as between speakers is expected.

Research questions 2 and 3 aim to capture this expected variation both on speaker as well as on item level. On item level, it will be attempted to identify patterns of target verbs. That is, depending on the results, verbs may be grouped depending on how acceptable they are in their split or non-split versions. In a next step, it can then be explored whether any of the factors recorded in the literature such as verbs' internal morphology, the semantic relationship between base or their frequency appear

to play a role in this grouping. On speaker level, it will be checked if the demographic variables have an effect on speakers' judgements and also more generally what overall rating patterns emerge between participants.

Concrete steps for this study will be to first identify suitable verbs. These are verbs which 1) are of the internal structure N + V and 2) whose finite form in non-V2 contexts is acceptable thus making them proper verbs. From there, those verbs' acceptability will be examined in V2 contexts by comparing both their acceptability in split and non-split contexts as well as contrasting their behaviour in those contexts with that of canonical particle verbs. This will result in the most extensive overview of these verbs in actual speaker data to date which can then be used to inform theory rather than the other way around.

5.1 Verb Selection

In previous experimental studies, target items have typically been chosen based on their prominence in theoretical literature. This approach presents two key issues: first, it remains unclear how theoretical authors originally selected these verbs. Second, the relevance of these items to actual language use is uncertain—specifically, whether speakers use them at all and whether their usage patterns align with syntacticians' judgments. To tackle these concerns, the present study employed a three-step selection process: (a) identifying verbs from the literature, (b) verifying their corpus frequency, and (c) assessing their usability as canonical verbs in a pretest. This procedure, as schematized in Figure 1, ensures that the selected verbs are not only recognized as relevant by syntactic theorists but are also frequent enough and widely acknowledged as actual verbs by speakers.

Forche (2020) conducted a literature review of potential V2 verbs and presents a large list of 1791 verbs which have been claimed to be non-V2 in the appendix of his work. He includes all verbs that at least one publication has treated as non-V2, even if other authors disagree with this judgement. Out of these, 638 are constructions consisting of a verb preceded by a nominal element and are therefore of interest for the present study. Items were excluded from the final selection when either the nominal or the verbal component cannot stand on its own either because it does not exist as an independent lexical item (e.g. *eifersüchteln* 'be jealous' where *süchteln* is not a word on its own*)* or because the nominal part contained a linking phoneme (e.g. *podiumsdiskutieren* 'podium.discuss'). Since the main experiment intended to test target items in both their non-split as well as their split form, these verbs were excluded. Furthermore, verbs whose base verb already contained a particle (e.g. *notumsiedeln* 'emergency.relocate') were excluded because it has been suggested that such double complexity may lead to structural conflicts responsible for verbs' inability to appear in V2 (Freywald & Simon, 2007). These criteria resulted in a final list of 495 verbs which were searched for in the corpus analysis.

From a look at the list of verbs treated in the literature, it immediately becomes clear that a high number of them describe highly specific activities – often specific to certain professions such as *punktschweißen* 'point.weld' or *ölhärten* 'oil.harden'. Irrespective of their status and usability as verbs which will be discussed below, this warrants the question of the general frequency of these items. As it has been suggested that potential non-V2 verbs might only be unacceptable due to dysfunctional inflectional paradigms which in turn are caused by verbs' low frequency. Items which are very infrequent, or which do not exist in speakers' mental lexicon at all will likely receive lower ratings in the experiment. It would then be unclear whether lower ratings stem from items' syntactic unacceptability or from their infrequency. This is why a corpus search of all 495 preselected items from Forche (2020) was performed with the goal of only selecting the most frequent items for the

experiment in the Deutscher Referenzkorpus DeReKo (Institut für Deutsche Sprache, 2023). This corpus is maintained by the Leibniz-Institut für Deutsche Sprache and contains 57 billion words from German newspaper, Wikipedia and online forum texts from the years 1999 to 2022. The complete corpus was searched for the infinitival form of all selected verbs. Items' absolute frequency was recorded which resulted in a ranking of the most frequent verbs.

There are a number of methodological issues in studying corpus frequency in general and particle verbs' frequency in particular: first, a general concern is that the corpus used is one of written texts only. As discussed before, however, N + V constructions may often be a product of spontaneous speech in spoken contexts. It is therefore likely that they are not very frequent in revised written texts (Forche, 2020) such as the newspaper articles in the corpus.

A practical issue with the corpus search is that the function to search by word class in a tree tagged grammar is not always reliable. This is because in many cases items' surface form may be used both as a verb and as a noun which in turn means that it is not always clear if the recorded frequency is a measurement of the item's occurrence as a *verb*. An example of this is *abendessen* which can mean both 'dinner' and '(to) evening eat'.

Finally, the most important problem specific to (potential) particle verbs is their (potential) separability: the corpus does not include any reliable and exhaustive way of recording the frequency of particle verbs in their *split* occurrence as those are not recognised as *one* word. Searching for this by hand by, for example, extracting items with an additional script, would not solve the problem. This is because in V2 contexts other constituents such as arguments and adjuncts can occur between base and particle which means that an unknown number of words may intervene between the two parts of the verb. This issue is discussed by Czypionka et al. (2019) who argue that particle verbs' frequencies must be 'systematically underestimated' (ibid.: 327) as a result of this.

The issues discussed here demonstrate that the results of the corpus search can only be seen as an approximation of items' true frequency. Despite these difficulties in determining frequency, the corpus search nevertheless helped in filtering the data: 191 verbs did not occur in the corpus at all. Hence, instead of trusting the measure of what the most frequent verbs are, this corpus research should rather be seen as a means to eliminate verbs from the final selection. Its main purpose was therefore to filter out and decrease the initial selection since that is too large to be fully tested in a behavioural study. The 60 items with the highest frequency were selected for the pretest.

5.2 Pretest

As the corpus study had identified items that occur as lexical items, the purpose of the pretest was to identify those which could be used as *verbs*. The 60 items selected in the corpus search were therefore presented in clause-final position in a subclause embedded under either a verb of cognition (*denken* 'think', *wissen* 'know', *vermuten* 'assume') or of perception (*sehen* 'see' *hören* 'hear', *beobachten* 'observe') which fitted the overall context as in (22). In addition to the 60 target items, 40 control items were included which were of the same structure as (22). As the purpose of the pretest was to identify the highest rated target verbs, the inclusion of control items was not strictly necessary. However, they were still added to the pretest to increase variability and prevent participants from only encountering the same type of items.

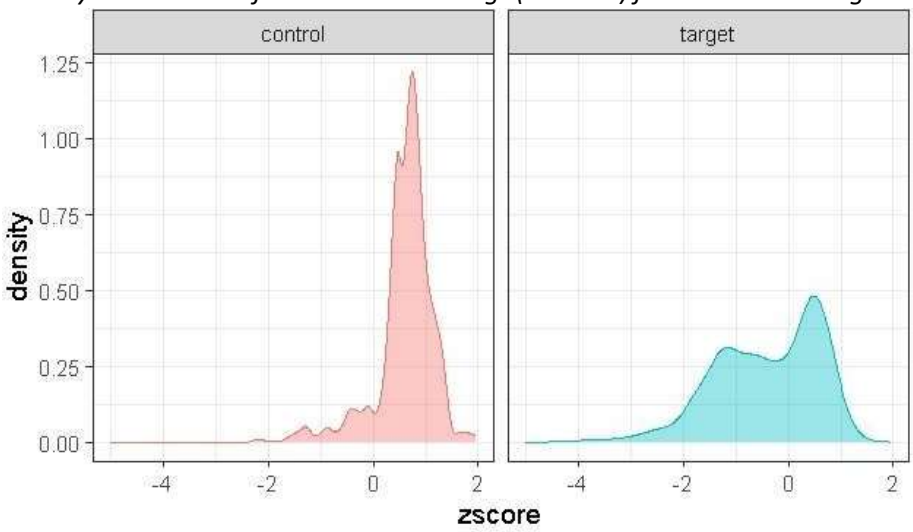
(22) Sabine sieht, dass der Künstler bauchspricht.
Sabine see.3SG that the artist belly.speak.3SG
'Sabine sees that the artist belly speaks.'

Sentences were presented in written form and participants rated their naturalness on a scale from 0 to 100 while being instructed to imagine the sentence being uttered by a native speaker of German. 62 native speakers of German recruited through personal communication and on social media participated in the experiment. Items were distributed across three experimental lists which were randomly assigned to participants so that each target verb was judged by 20 to 30 speakers. As the pretest only served the purpose of validating the selection of items thus far, no demographic data about participants was collected.

A linear mixed effects model with the outcome variable z-score and the predictor VERB TYPE which included participant and item as random effects revealed that ratings standardised per participant for target verbs (M = - 0.4; SD = 1) differed from those of control items (M = 0.59; SD = 0.55) in having significantly lower acceptability scores (β = -1.01, SE_β = 0.146, p < 0.001). This can also be visually observed in Figure 2 which depicts the density of the frequency of standardised acceptability ratings. The figure also shows the difference in variance between target and control condition: results are much more spread out for the target verbs while controls have a clear peak. This suggests greater variability in how participants perceived the acceptability of target verbs. When claiming that the verbs under examination are unacceptable in V2 contexts, authors have standardly assumed that they are perfectly acceptable in non-V2 positions. However, the data obtained here suggests that this is not the case. In fact, the obtained data leaves the possibility open that items may not be acceptable in V2 because they are either not acceptable as verbs at all or because they are highly infrequent. This result is therefore important to keep in mind when examining the same verbs in V2 contexts in the main experiment and drawing conclusions about their properties.

Figure 2

Density distribution of standardised ratings (z-scores) for control and target verbs in the pretest



After the pretest, the 30 most highly rated items based on raw scores were selected to be used in the main experiment. These verbs had received a mean standardised rating of 0.05 (*SD* = 0.74). Having done as much as possible to assure the selection of verbs that a) exist and b) are more likely to be used as verbs, the main experiment tested these verbs' acceptability in V2 contexts while taking their perceived frequency and speakers' language background into account.

6 Main Experiment

6.1 Methodology

In the main experiment, the 30 items selected from the pretest were tested in different syntactic conditions and contrasted with regular German particle and prefix verbs in a web-based acceptability judgment task of spoken recordings.

6.2 Participants

Participants were recruited through personal communication and on social media. Native speakers of German with no diagnosed hearing or language disorders participated. Upon entering the experiment, participants had to confirm they were in quiet surroundings and were able to listen to audio recordings on their device. To confirm their audio was indeed working a short audio recording describing holiday plans was presented before the start of the experiment. Participants were then asked to select what the recording had been about out of a list of four options. All participants answered this question correctly. 141 participants completed the survey. Their mean age was 42 (*SD* = 14.86; Range: 16 – 75). 91 were female, 41 male and five of third gender. Four other participants did not disclose their gender. 132 participants grew up in Germany, four in Switzerland and Austria each and one in Luxembourg.

6.3 Materials

The 30 verbs selected from the pretest and their translations are listed in Table 2. Verbs were grouped into six experimental conditions crossing the factors VERB TYPE and CONTEXT. For the first factor, complex verbs of the structure N + V as selected in the pretest, were contrasted with verbs whose status as either separable particle or inseparable prefix verbs is undebated. Since many of the target verbs refer to specific actions not likely to be performed by just anyone, those verbs were first matched with a semantically appropriate animate agent that was then used as the subject in its definite singular form. Nouns were counterbalanced for male and female referents. In a second step, canonical particle and prefix verbs were selected that also matched the chosen subjects. Finally, for all three of these conditions split and non-split structures were created. This resulted in the experimental conditions shown in Table 3. The full list of experimental stimuli is given in Appendix B.

Table 2

List of N + V verbs used in the main experiment

Verb	Gloss	Translation
bauchreden	belly.speak	(to) ventriloquize
bauchtanzen	belly.dance	(to) belly dance
bergsteigen	mountain.climb	(to) mountaineer
bettnässen	bed.wet	(to) bed wet
bogenschießen	arrow.shoot	(to) arrow shoot
bruchlanden	break.land	(to) make a crash landing

brustschwimmen	breast.swim	(to) swim breast strokes
dampfgaren	steam.cook	(to) steam cook
dauerparken	duration.park	(to) park permanently
ehebrechen	marriage.break	(to) commit adultery
eiskunstlaufen	ice.art.walk	(to) figure skate
fallschirmspringen	parachute.jump	(to) parachute
fotokopieren	photo.copy	(to) photocopy
heimwerken	home.work	(to) do do-it-yourself crafting
kopfrechnen	head.calculate	(to) calculate in one's head
korrekturlesen	correction.read	(to) proofread
maßschneidern	measurement.tailor	(to) custom tailor
notlanden	emergency.land	(to) emergency land
notoperieren	emergency.operate	(to) perform emergency surgery
probefahren	trial.drive	} (to) drive/walk/sit/live on trial basis
probelaufen	trial.walk	
probesitzen	trial.sit	
probewohnen	trial.live	
raubkopieren	theft.copy	(to) pirate
schlafwandeln	sleep.wander	(to) sleepwalk
seilspringen	rope.jump	(to) do rope skipping
seiltanzen	rope.dance	(to) do rope dancing
skispringen	ski.jump	(to) do ski jumping
stoßlüften	hit.ventilate	(to) ventilate
tagträumen	day.dream	(to) daydream

Table 3

Example of experimental conditions used in the main experiment.

Context/ Verb Type	**Target**	**Particle**	**Prefix**
Split	Die Frau bauchtanzt oft.	Die Frau einkauft oft.	Die Frau verkauft das Buch.
Non-split	Die Frau tanzt oft bauch. 'The woman often belly dances.'	Die Frau kauft oft ein. 'The woman often goes shopping.'	Die Frau kauft das Buch ver. 'The woman sells the book.'

A challenge arising from the inclusion of regular prefix verbs that those are predominately transitive which means that using them intransitively would likely decrease acceptability ratings. Therefore, those verbs were presented with a semantically fitting object which was singular, definite and inanimate in all cases. In all but two cases where dative was required by the verb, those were accusative objects. In contrast, in the original selection, target verbs were picked so that an intransitive reading would be available, in order not to interfere with a possible object reading of the nominal component in the split condition. This therefore creates an imbalance between conditions. As this condition only serves as an ungrammatical baseline test case for the separation of complex verbs, this difference is not as problematic. What had to be controlled for, however, was the linear distance between the base verb and the particle. To ensure this, a temporal adverb was placed in the object position in both the target and particle conditions. The adverb considered most semantically appropriate for the specific action was selected from *wieder* 'again', *noch* 'still', *oft* 'often', *manchmal* 'sometimes' and *regelmäßig* 'regularly', while maintaining a balanced distribution so that each adverb

appeared seven or eight times. As in the pretest, all verbs were inflected for third person present tense. Nominal particle-like elements and regular preposition-like particles would ideally have been paired with the same base verb, in order to control for base verb effects. In spite of this preference, this ideal is only possible for some of the selected verbs (e.g. *bergsteigen* 'mountain.climb'*; aussteigen* 'disembark'*; besteigen* 'mount'), and not productive in most cases.

Experimental lists were created by first distributing the non-target conditions across four lists using a Latin Square design, resulting in 15 items of each type per list. Due to the expected variation between both participants and items, it was considered to be most valuable and informative for present purposes to let participants view a higher number of target verbs. Hence, target verbs were only split into two lists both of which contained half of the items in their split and the other half in their non-split condition. As a next step, these two lists were matched to the remaining four lists by ensuring that participants would never see the same syntactic context from one set. This means that if the particle/prefix verb for a given set was presented in its split version, the target verb would be presented non-split and vice versa.

In addition to the experimental conditions, 45 filler items of varying degrees of acceptability were used. Out of those, ten were grammatical fillers of different syntactic structures. Fifteen were ungrammatical fillers out of which ten were rendered ungrammatical due to agreement violations while five others were word salads, i.e. non-sensical strings of random words. Grammatical and ungrammatical fillers were kept short and simple in structure because target sentences were short too. An additional twenty fillers were different types of grammatical norm violations. These included the use of *wie* as a comparative particle, the verb *tun '*(to) do' as an auxiliary, the progressive form, the use of dative case marking in prescriptively genitive environments and V2 word order in embedded clauses introduced by *weil* 'because'. The same fillers were used for all four lists resulting in a total of 105 items per list. 20 randomised versions of each list were created using the Mix software for randomisation (van Casteren & Davis, 2006) which resulted in 80 lists in total. Requirements for randomisation were that items from the same set should never directly follow each other and that no more than three target verbs should occur in immediate succession.

All experimental items were recorded by the same female native speaker of Standard German. Words where her pronunciation might deviate from the standard for regional or idiosyncratic reasons were not used. This did not affect the selection of verbs. In the target condition, the word accent was placed on the nominal particle as in *Der Wanderer bérgsteigt* 'The hiker mountaineers'. The only exception was the verb *fotokopieren* '(to) photo copy' which was stressed as *Der Direktor fotokopiért* 'The director photocopies' because the morpheme of Romance origin *-ieren* always bears the word accent (Hüning, 2018) and this is the pronunciation given by the German reference grammar Duden (Dudenredaktion, n.d.).

6.4 Perceived Frequency

As the corpus study revealed that many of the selected verbs are of low frequency and the pretest showed overall lower ratings for target verbs than for canonical items, perceived frequency was chosen as a measure in the main experiment. This was done to examine whether listeners would rate target items they thought appeared more frequently more highly. Materials for this experimental component were audio recordings of only the infinitival form of the target verbs spoken by the same speaker as in the main experiment.

6.5 Procedure

The experiment was conducted online using Qualtrics (Qualtrics, Provo, UT, January 2025). After a sound check and demographic questions, participants were presented with one of the 80 randomised lists of items. Participants were instructed to imagine themselves in a conversation with the speaker of the recording who they were told is a native speaker of German. They were asked to judge how natural the sentence they heard is on a sliding scale from 0 to 100 which was initially set to 50. Participants were able to see the value they selected on the scale. They were instructed to judge intuitively and not based on pronunciation. After they had rated all sentences, they were asked to guess what the experiment had been about.

Finally, in order to assess the effect of frequency on ratings, participants were presented with 15 randomly selected recordings of the infinitival form of the target verbs in isolation. They were asked the question *Wie häufig kommt dieses Wort im Deutschen vor?* 'How frequent is this word in German?' and rated its frequency on a one-hundred-point scale ranging from *sehr selten*, 'very infrequently' to *sehr häufig* 'very frequently'. Participants were explicitly instructed not to judge based on their personal usage of the target word but instead based on how frequently they estimate to encounter the word being used.

7 Analysis

7.1 Data Exclusion

Before statistical evaluation, exclusion criteria on participant level were checked. Two participants correctly guessed that the purpose of the study was to examine whether verbs with a nominal component behave more like particle or prefix verbs and were therefore excluded from the analysis. To ensure serious adherence to the task, word salad items functioned as an exclusion criterion: participants who rated more than two out the five word salad items as higher than their personal lowest score or as higher than 20 were excluded. This resulted in the exclusion of three participants. The results reported here are thus based on 136 participants. Finally, due to an error in how items were distributed across list, some participants saw the same verb both in its split and non-split form which was not intended by the original distribution criteria. To handle this issue, the second judgement of these verbs was excluded.

7.2 Statistical Analysis

7.2.1 Linear Mixed Models

All data analysis was carried out in R (version 4.3.1; R Core Team, 2025). Prior to analysis, acceptability scores and perceived frequency ratings were standardised to z-scores for each participant over all items to control for differential use of the rating scale. In order to examine the effect of verb type and context on acceptability score, two linear mixed effects models were fitted in R using the *lmer* function from the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). The first model included z-scores of all three verb types, while the second model only used target verbs' ratings as the outcome variable.

In order to answer RQ1, Model 1 was used to compare ratings across the three verb types and therefore contained VERB and CONTEXT as well as their interaction as fixed effects. Both variables were coded using sum contrasts. The model's random effects structure was fitted using the Akaike Information

Criterion (AIC). Appendix C contains an overview of all tested models. First, only a by-item random intercept was included, and the model was then improved and extended stepwise. It was found that the model with minimal AIC and thus highest explanatory value included by-item as well as by-participant random slopes. Results of the assumption checking performed as part of model criticism are given in Appendix D. Following the evaluation of the model results, all experimental conditions were also compared pairwise using estimated marginal means extracted with the *emmeans* package in R (Lenth, 2023).

Model 2 was used to identify factors specifically affecting the ratings of target verbs on both item and participant level and thus aimed to answer RQs 2 and 3. It included standardised ratings of target verbs as the outcome variable. The factor CONTEXT was sum coded again. Besides this, the model included the participant-level fixed effect AGE. Because more than a quarter of participants (*n* = 37) had chosen not to disclose their level of education, this variable was not included as a predictor. Additionally, despite the large number of participants, there was no balance in their regions of origin, so no regional variable was included either. On item level, the fixed effect FREQUENCY represented the mean standardised perceived frequency rating per item. Raw frequency ratings per participant per item could not be used since each participant only judged the frequency of fifteen out of thirty verbs. Finally, the factor PRETEST was included in form of the standardised average pretest score per target verb.

First, the maximal model containing all of the above variables and their interactions was tested which was subsequently simplified with AIC. Next, the random effects structure was fitted. In addition to the above-described fixed effects, the final model contained the interactions of CONTEXT x PRETEST, FREQUENCY x PRETEST and the three-way interaction CONTEXT x PRETEST x FREQUENCY as well as random slopes and intercepts for item and participant. From this second model, random effects were extracted and visualised. AIC results as well as model criticism can be found in the Appendices B and C. Testing for homogeneity of variance, it was found that there was a correlation between the fitted values and the size of the error. This points towards an unequal distribution of variability in responses which will be discussed further in the subsequent sections.

7.2.2 Cluster Analysis

While linear mixed effects models provide an indication of the general tendency of ratings of the target verbs overall, their results do not straightforwardly present insights into potential patterns in target verbs' scores or participants' rating behaviour. The random slopes for both items and participants added explanatory power to the models which is in line with the expected variation on item and participant level. A more fine-grained analysis of this variation was therefore deemed useful in order to answer RQs 2 and 3. This was achieved by conducting a hierarchical and k-means cluster analysis for both target verbs and participants with the aim of detecting whether there are patterns in the observed variation.

Following Levshina (2015), a distance matrix of the Euclidean distances between each item's acceptability vector was computed. From this, verbs were clustered hierarchically based on their distance scores using Ward's minimum variance method. To determine which verbs' behaviour could be combined into groups, the ideal number of clusters was determined using the Silhouette function from the *cluster* package in R (Maechler et al., 2025). This function computes the average silhouette width for each number of clusters which reflects how well each verb fits within its assigned cluster compared to other clusters. Silhouette scores range from 0 to 1 with higher values indicating clearer groupings.

This method revealed the optimal number of clusters in the data to be six with an average Silhouette width of 0.44. Verbs' cluster membership was then determined using k-means clustering. This algorithm first randomly assigns items to one of the k clusters (six in this case), calculates the mean of each cluster and then reassigns the items to the cluster with the nearest mean. This process is repeated until clusters do not change anymore.

This same procedure was repeated for participants to extract patterns in rating behaviour. Theoretically, it would be possible for each participant to represent their own cluster. Since a large number of clusters would not be informative, however, only Silhouette scores for clusters where $k < 10$ were computed. Here, the optimal number of clusters was revealed to be nine with a Silhouette score of 0.39. Silhouette scores for all numbers of clusters for both items and participants can be found in Appendix D.

8 Results

8.1 Descriptives

Table 4 shows the average standardised and unstandardised ratings for each of the six experimental conditions. These results are also visualised in the density plots in Figure 3 for particle and prefix verbs and Figure 4 for target verbs. Both the numerical data as well as the visualisation are indicative of a clear difference between the three conditions: while ratings of both regular prefix and particle verbs reflect a clear preference for one of the syntactic contexts, target verbs' ratings are more spread out with average ratings around the centre of the scale for both contexts. Additionally, target verbs display a bimodal distribution in their split form which suggest that this context received above and below average ratings to similar extents. In contrast, ratings for target verbs' non-split forms are more skewed towards the left.

Results of filler items are presented in Table 5 as well as visualised in Figure 5. These show that, as expected, grammatical (M = 1.20, SD = 0.33) and ungrammatical (M = -0.86, SD = 0.46) fillers received overall high and low ratings respectively, while the norm violation items reflect more disagreement in judgements (M = 0.27, SD = 0.72). More detailed results for each type of norm violation are given in Appendix C.

Table 4

Summary of mean descriptive results per experimental condition. Standard variations are given in parentheses. "Overall" indicates the overall mean per condition.

Verb	**Split**		**Non-Split**		**Overall**	
	Standard.	**Unstandard.**	**Standard.**	**Unstandard.**	**Standard.**	**Unstandard.**
Prefix	-0.99 (0.45)	11.33 (18.4)	1.06 (0.49)	91.42 (19.89)	0.03 (1.13)	51.39 (44.39)
Particle	1.09 (0.4)	92.59 (16.54)	-0.96 (0.32)	12.51 (15.59)	0.06 (1.08)	52.63 (43.16)
Target	0.01 (0.85)	50.61 (36.65)	-0.25 (0.78)	40.31 (34.02)	-0.12 (0.83)	45.26 (35.68)
Overall	0.03 (1.001)	51.33 (40.69)	-0.104 (0.96)	46.04 (0.96)		

Figure 3

Density plots of z-score distribution of particle and prefix verbs in both experimental contexts

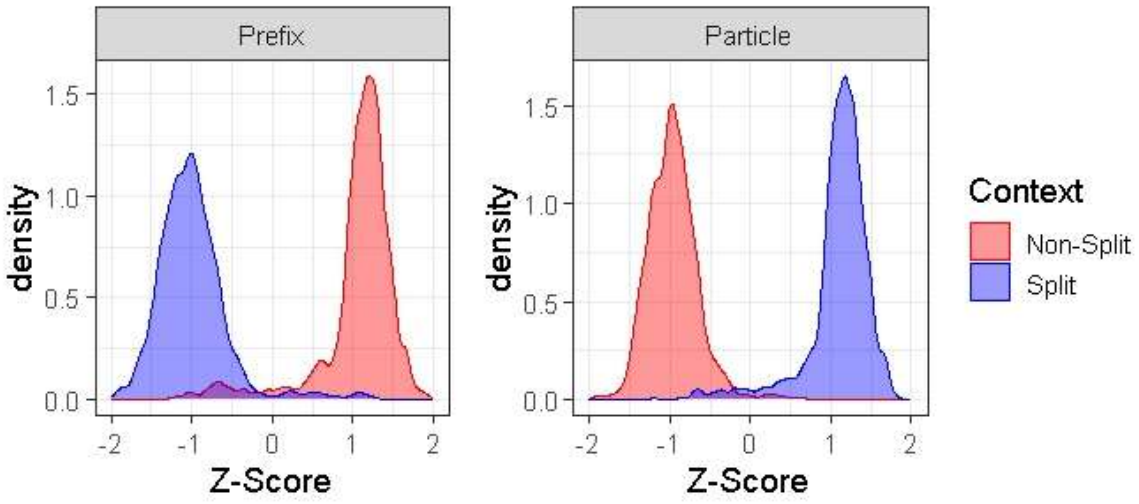


Figure 4

Density plots of z-score distribution of target verbs in both syntactic contexts

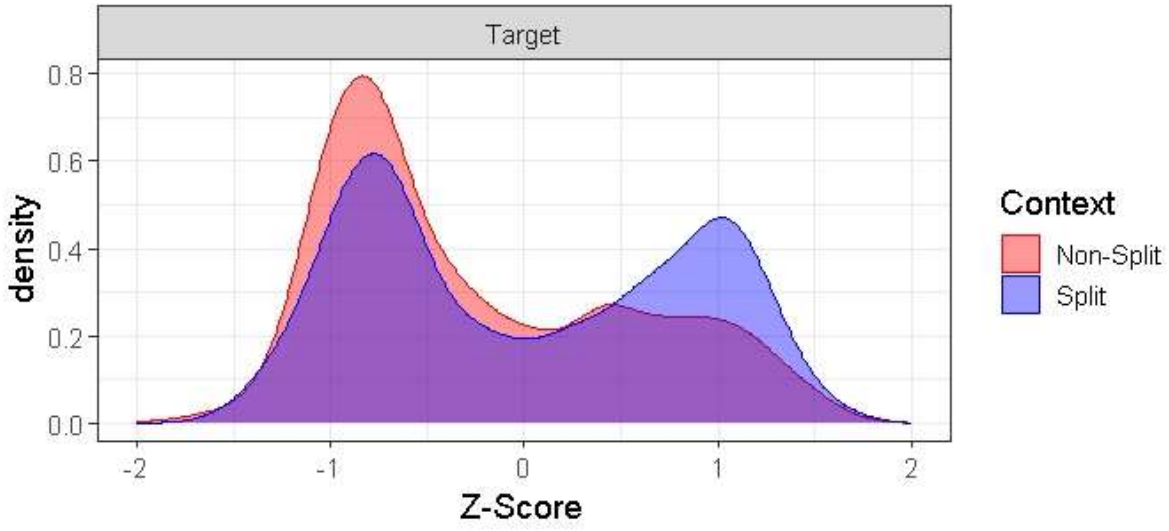


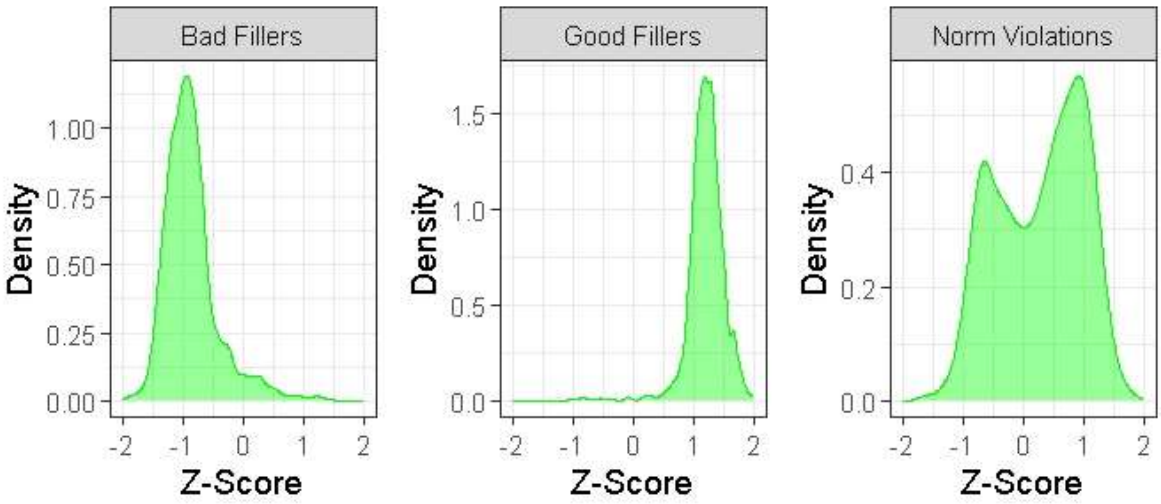
Table 5

Summary of ratings of standardised and unstandardised ratings of filler items. Standard deviations are given in parentheses

	Standardised	Unstandardised
Grammatical Fillers	1.20 (0.33)	96.48 (11.42)
Ungrammatical Fillers	-0.86 (0.46)	16.33 (20.86)
Norm Violations	0.27 (0.72)	60.54 (32.37)

Figure 5

Density plots of z-score distribution of filler items



8.2 Model 1

Model 1 tested whether the observed descriptive differences between the three verb types were statistically significant. Results are summarised in Table 6. As the predictors were sum coded, the estimates reflect differences from the grand mean. The model revealed a significant difference between particle and target verbs (β = 0.07, SE_β = 0.03, 95% CI [0.01, 0.13], p = .023): target verbs (M = -0.12, SD = 0.83) received overall lower ratings than particle verbs (M = 0.06, SD = 1.08). The difference between prefix (M = 0.03, SD = 1.13) and target verbs, however, was not significant (β = 0.04, SE_β = 0.03, 95% CI [−0.02, 0.10], p = .207). The model also did not reveal a significant main effect of CONTEXT: ratings in split contexts (M = 0.03, SD = 1.001) were slightly higher than those in non-split contexts (M = –0.10, SD = 0.96), but this difference was not significant (β = −0.04, SE_β = 0.03, 95% CI [−0.11, 0.02], p = .174).

Table 6

Summary of Model 1 results with the outcome variable z-score rating of all experimental conditions.

Fixed Effects						
Predictors	*β*	95% CI	SE	df	*t*	*p*
Intercept	−0.00	[−0.05, 0.04]	0.02	109.40	−0.16	.874
Verb (particle vs. target)	0.07	[0.01, 0.13]	0.03	92.57	2.31	.023 *
Verb (prefix vs. target)	0.04	[−0.02, 0.10]	0.03	92.58	1.27	.207
Context (split vs. non-split)	−0.04	[−0.11, 0.02]	0.03	97.79	−1.37	.174
Verb × Context (particle)	−0.98	[−1.07, −0.91]	0.04	89.74	−22.59	< .001 ***
Verb × Context (prefix)	1.07	[0.99, 1.16]	0.04	89.75	24.54	< .001 ***

Random Effects			
Group	Name	Variance	SD
ResponseId	Intercept	0.006	0.08
ResponseId	Context split	0.006	0.08
Word	Intercept	0.035	0.19
Word	Context split	0.082	0.29
Residual		0.257	0.51

Notably, a significant VERB × CONTEXT interaction emerged. In particular, prefix verbs were rated low in split contexts (*M* = –0.99, *SD* = 0.45) and highly in non-split contexts (*M* = 1.06, *SD* = 0.49), whereas the opposite pattern was found for particle verbs, which were rated highly in split contexts (*M* = 1.09, *SD* = 0.40) and low in non-split contexts (*M* = –0.96, *SD* = 0.32). VERB x CONTEXT interaction effects were significant for both comparisons. Expectedly, the interaction between context and particle verbs was negative (β = −0.98, SE_β = 0.04, 95% CI [−1.07, −0.91], p < .001), while the interaction between context and prefix verbs was positive (β = 1.07, SE_β = 0.04, 95% CI [0.99, 1.16], p < .001).

Delving deeper in these results, the pairwise estimated marginal means comparisons summarised in Table 7 additionally showed that target verbs' ratings in both contexts differed significantly from all conditions of control verbs (all p < .001). However, target verb ratings did not differ from each other in the two syntactic configurations (β = –0.26, SE_β = 0.11, p = 0.15). All in all, target verbs consistently received intermediate ratings across both conditions (Split: *M* = 0.01, *SD* = 0.8); Non-Split: *M* = -0.25, *SD* = 0.78) which were lower than the two types of canonical verbs in their respective preferred configuration but significantly higher than those verbs in their dispreferred forms.

Table 7

Results of the estimated marginal means comparison between all experimental conditions

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
target non-split – particle non-split	0.73	0.09	8.43	< .0001 ***
target non-split – prefix non-split	–1.29	0.09	–14.95	< .0001 ***
target non-split – target split	–0.26	0.17	–2.41	.15
target non-split – particle split	–1.32	0.09	–14.52	< .0001 ***
target non-split – prefix split	0.76	0.09	8.33	< .0001 ***
particle non-split – prefix non-split	–2.02	0.09	–23.18	< .0001 ***
particle non-split – target split	–0.99	0.09	–10.81	< .0001 ***
particle non-split – particle split	–2.05	0.12	–19.08	< .0001 ***

particle non-split – prefix split	0.03	0.09	0.32	.99
prefix non-split – target split	1.04	0.09	11.37	< .0001 ***
prefix non-split – particle split	–0.02	0.09	–0.32	.99
prefix non-split – prefix split	2.05	0.12	19.08	< .0001***
target split – particle split	–1.07	0.09	–11.4	< .0001***
target split – prefix split	1.02	0.09	10.86	< .0001***
particle split – prefix split	2.08	0.09	22.12	< .0001***

8.3 Model 2

Having confirmed that the target verbs do not behave like either of the other two verb types, Model 2 aimed to explore this behaviour in more depth. Its results, including all predictors, are shown in Table 8. As already indicated by the pairwise comparisons following the results of Model 1, the two syntactic contexts did not significantly differ for target verbs. This is also reflected in the current model by the absence of a significant main effect of CONTEXT (β = –0.07, SE_β = 0.10, 95% CI [–0.28, 0.13], p = .47), indicating that target verbs received similar ratings in both split and non-split contexts.

Table 8

Summary of Model 2 results with the outcome variable z-score of target verbs

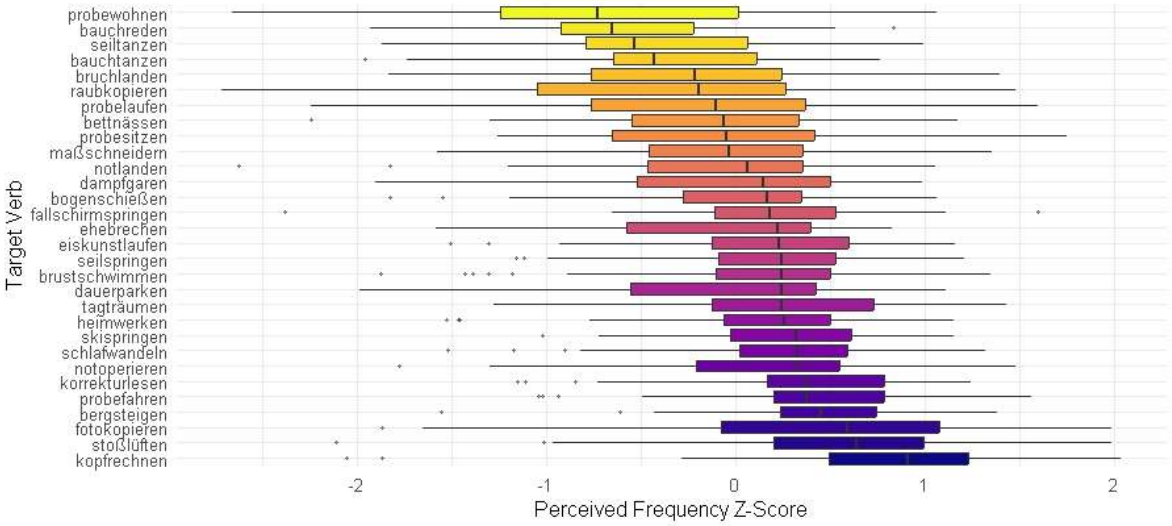
Fixed Effects						
Predictors	β	[95% CI]	SE	df	***t***	***p***
Intercept	–0.03	[–0.18, 0.11]	0.08	86.97	–0.46	.65
Context	–0.07	[–0.28, 0.13]	0.1	26.88	–0.73	.47
Pretest	0.45	[0.07, 0.84]	0.19	25.98	2.46	.021 *
Frequency	0.09	[–0.26, 0.44]	0.17	25.99	0.53	.60
Age	–0.0032	[–0.0057, –0.0008]	0.0012	134.42	–2.63	.009 **
Context × Pretest	–0.39	[–1.10, 0.31]	0.34	25.92	–1.15	.26
Context × Frequency	0.16	[–0.48, 0.81]	0.32	25.93	0.52	.61
Pretest × Frequency	0.71	[–0.28, 1.71]	0.48	26.40	1.48	.15
Context × Pretest × Frequency	–0.87	[–2.74, 0.99]	0.9	26.10	–0.97	.34

Random Effects			
Group	Name	Variance	SD
Participant	Intercept	0.032	0.18
Participant	Context split	0.094	0.31
Item	Intercept	0.287	0.54
Item	Context split	0.899	0.95
Residual		0.324	0.57

Figure 6 displays a bar plot of the mean standardised perceived frequency estimates per verb (M = 0.05, SD = 0.33). These scores, however, did not significantly predict acceptability ratings (β = 0.09, $SE\beta$ = 0.17, 95% CI [–0.26, 0.44], p = .602) and the interactions between both FREQUENCY and PRETEST (β = 0.71, $SE\beta$ = 0.48, 95% CI [–0.28, 1.71], p = .152) and between FREQUENCY and CONTEXT (β = 0.16, $SE\beta$ = 0.32, 95% CI [–0.48, 0.81], p = .607) as well as the three-way interaction between FREQUENCY, PRETEST and CONTEXT (β = –0.87, $SE\beta$ = 0.90, 95% CI [–2.74, 0.99], p = .341) all remained insignificant. Overall, this means that ratings remained steady across varying levels of perceived frequency, and that this was the case in both syntactic contexts and irrespective of pretest score.

Figure 6

Z-scores of perceived frequency ratings per target verb



In contrast, the standardised PRETEST ratings of the verbs did have a marginally significant effect on ratings in the main experiment (β = 0.45, SE_β = 0.19, 95% CI [0.07, 0.84], p = .021), with higher pretest scores predicting higher acceptability. This effect did not differ depending on context, as the interaction between CONTEXT and PRETEST was not significant (β = –0.39, SE_β = 0.34, 95% CI [–1.10, 0.31], p = .263). The relationship between pretest scores and mean ratings is illustrated in Figure 7 for both contexts. As the pretest tested verbs' acceptability in clause-final (non-V2) contexts, this relationship indicates that verbs which scored higher in the pretest also received higher ratings in the main experiment.

On participant level, results furthermore revealed a significant main effect of AGE (β = –0.0031, SE = 0.0012, 95% CI [–0.0056, –0.0008], t = –2.63, p = .009), indicating that as age increased, acceptability ratings slightly decreased. This relationship is visualised in Figure 8 which shows each participant's mean target verb rating per context plotted against their age.

Figure 7

Mean standardised pretest results per verb plotted against mean standardised result in the main experiment

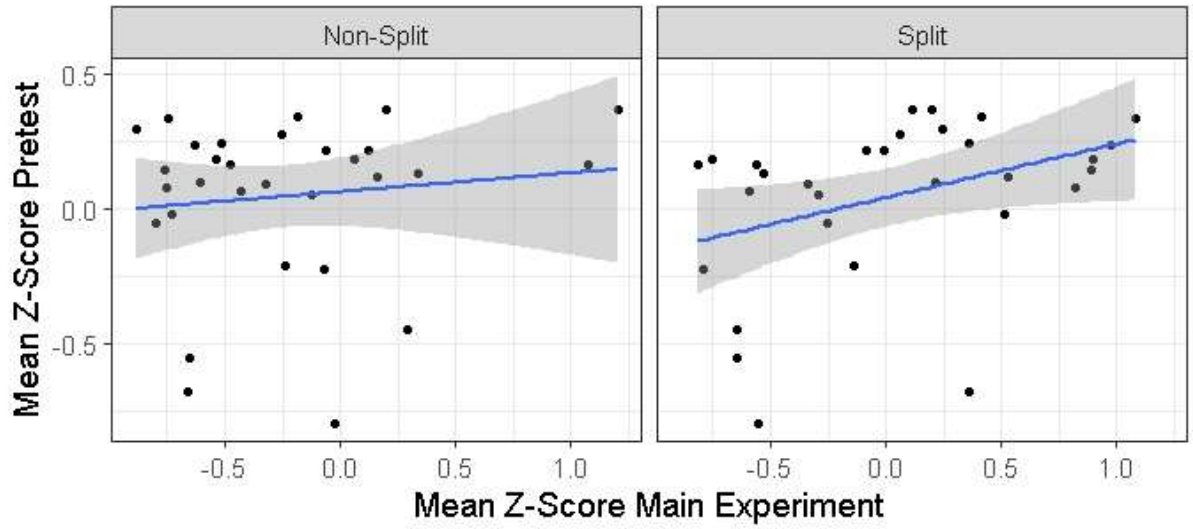
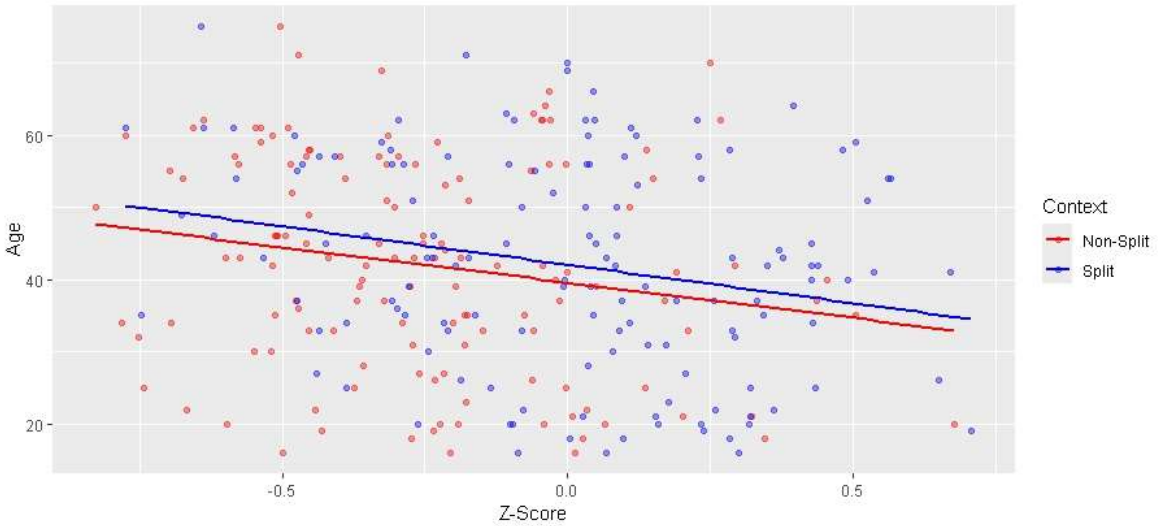


Figure 8

Mean standardised rating of target verbs per participant per context plotted against participants' age



Finally, since the model selection process showed an improved model fit when including random slopes those were extracted and are visualised for items in Figure 9. This visualisation shows how much each verb deviated from the mean of the level non-split. That is, positive effects indicate that a verb received higher than average ratings in the given context while negative effects indicate lower than average effects. To further explore this variation between items, Figure 10 can be inspected which depicts the mean standardised score per verb in each syntactic context compared to the average scores of prefix and particle verbs as well as the average score of target verbs themselves. It shows that verbs showed no uniform results in either context. In addition to the variation between verbs, there was also considerable variation in how different participants rated any given verb. This spread of ratings per verb is visualised in Figure 11 which shows that disagreement was greater for some than for other verbs.

Figure 9

Random slopes for participants for the context non-split extracted from Model 2

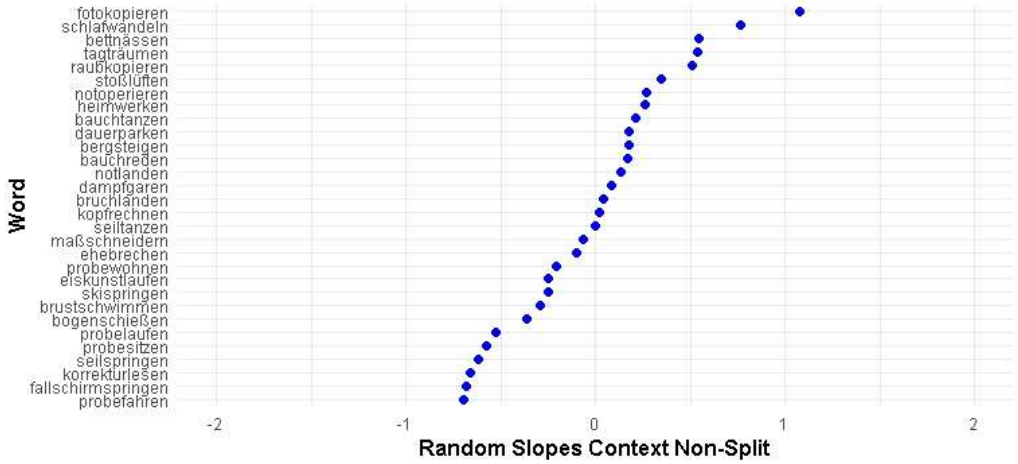


Figure 10

Bar plots of the mean z-score each target verb received in non-split and split contexts. Dotted lines represent the mean ratings of prefix, particle, and target verbs in each condition

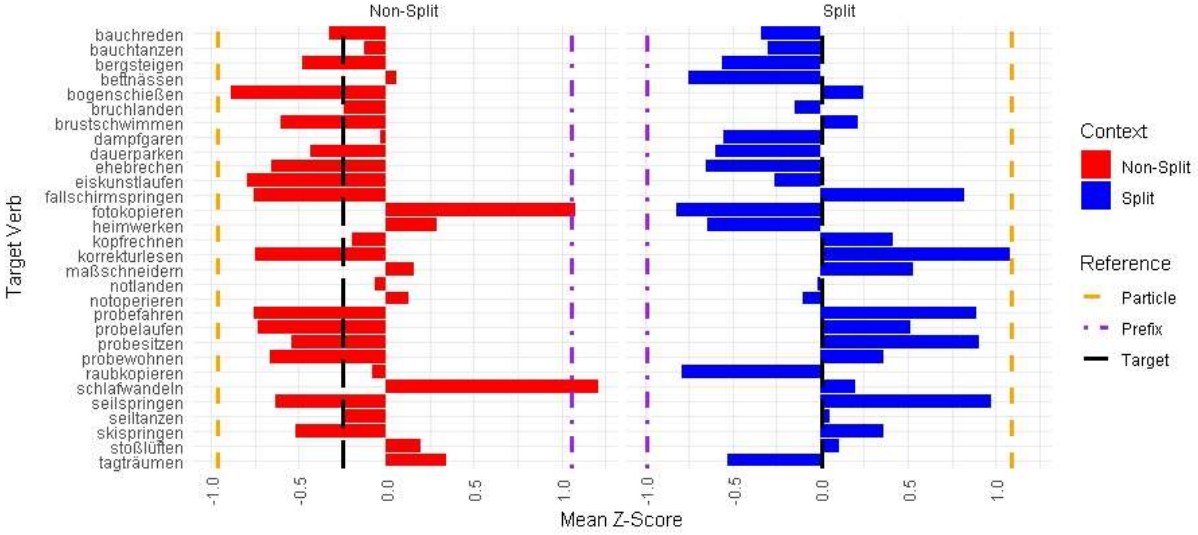
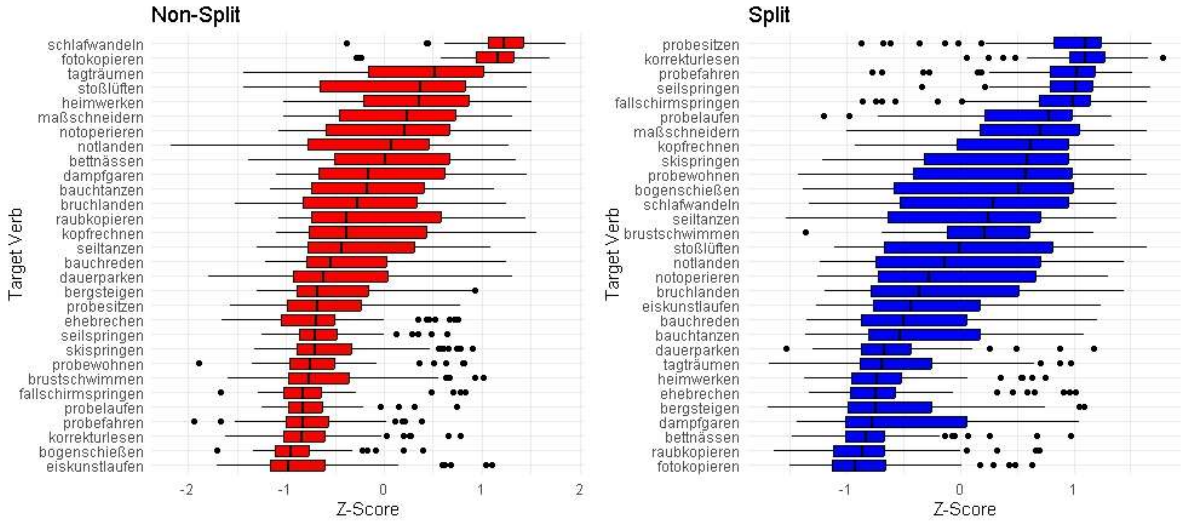


Figure 11

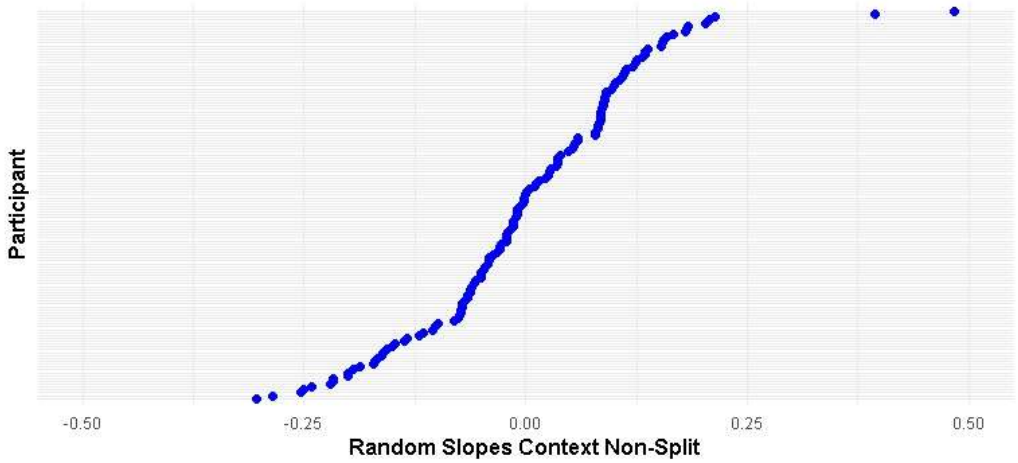
Boxplots of ratings per target verb in each context



Given that random slopes for participants added explanatory value to the models, they were also extracted for Model 2 and are presented in Figure 12 which again shows positive and negative effects. Both item as well as participant random effects suggest that there are patterns in rating behaviour which will be explored in more detail in the subsequent cluster analysis.

Figure 12

Random slopes for participants for the context non-split extracted from Model 2



8.4 Cluster Analysis

8.4.1 Verbs

The results of the hierarchical cluster analysis of verbs are displayed in the dendrogram in Figure 13. Verbs being adjacent each other signifies that these verbs have the smallest distance between their acceptability vectors which means that each verb shares the greatest similarity in ratings with its neighbours. The results of the subsequent k-means analysis are summarised in Table 9 which shows which verbs fall into the same cluster. Ratings of the final six clusters are summarised in Table 10 and visualised in the boxplots in Figure 14.

Figure 13

Dendrogram of hierarchical clustering of target verbs

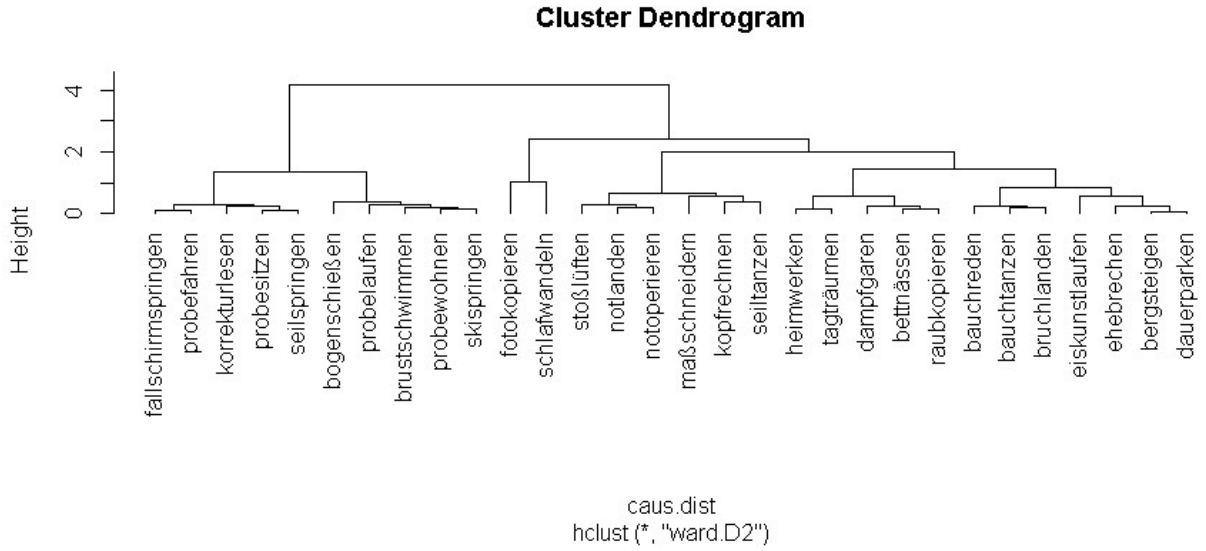


Table 9

List of target verbs and their cluster membership

Cluster	Verbs
1	fotokopieren, schlafwandeln
2	bettnässen, dampfgaren, heimwerken, raubkopieren, tagträumen
3	fallschirmspringen, korrekturlesen, probefahren, probesitzen, seilspringen
4	bogenschießen, brustschwimmen, probelaufen, probewohnen, skispringen
5	bruchlanden, kopfrechnen, maßschneidern, notlanden, notoperieren, seiltanzen, stoßlüften
6	bauchreden, bergsteigen, dauerparken, ehebrechen, eiskunstlaufen

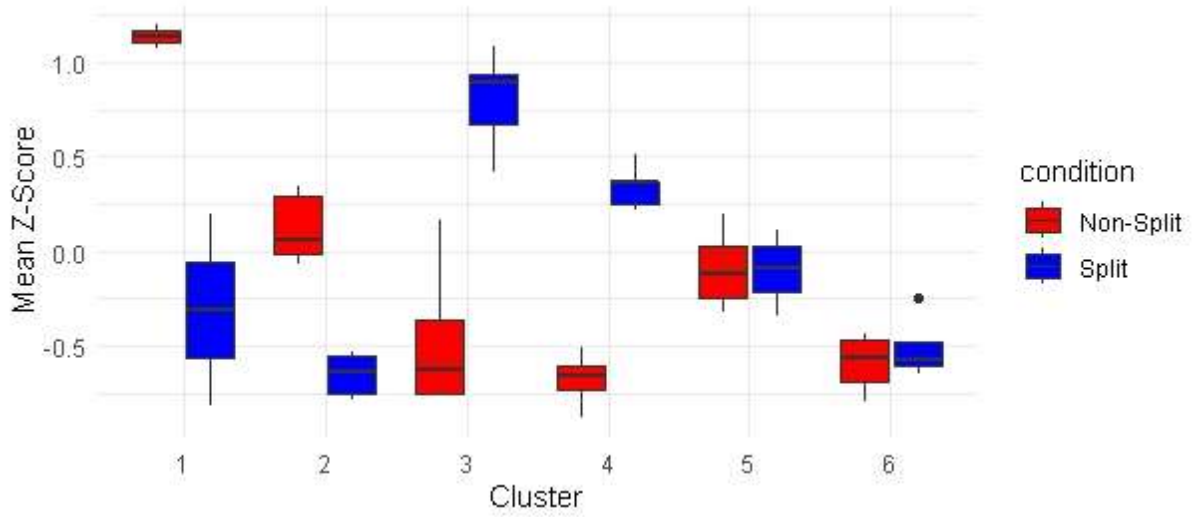
Table 10

Mean standardised scores in each condition per cluster of target items. Standard deviations are given in parentheses

Cluster	**1 (*n* =2)**	**2 (*n* = 6)**	**3 (*n* = 5)**	**4 (*n* = 5)**	**5 (*n* = 7)**	**6 (*n* = 5)**
Split	-0.31 (0.72)	-0.59 (0.18)	0.93 (0.1)	0.34 (0.12)	0.13 (0.25)	-0.48 (0.17)
Non-split	1.14 (0.09)	0.08 (0.19)	-0.69 (0.1)	-0.68 (0.14)	-0.04 (0.2)	-0.54 (0.19)

Figure 14

Boxplots of mean ratings per cluster of items in the two syntactic contexts



The six clusters can be broadly grouped into three classes: 1) verbs whose non-split form is preferred (Clusters 1 and 2), 2) verbs whose split form is preferred (Clusters 3 and 4) and 3) verbs that received similar ratings in both contexts (Clusters 5 and 6).

The two verbs in Cluster 1 are acceptable in their non-split (*M* = 1.14, *SD* = 0.09) but less so in their split form (*M* = -0.31, *SD* = 0.72). Results per verb in this cluster are summarised in Table 11 and visualised in Figure 15. Overall, these two verbs are the highest rated across the entire experiment.

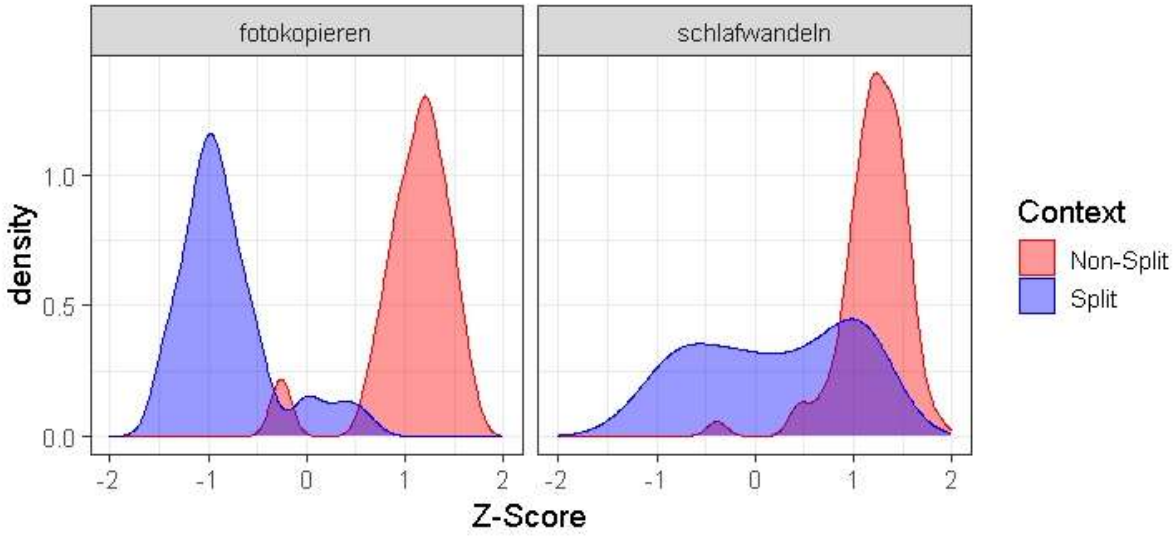
Table 11

Standardised and unstandardised means and standard deviations of verbs in Cluster 1. Standard deviations are given in parentheses

Verb	Split		Non-split	
	Standardised	**Unstandardised**	**Standardised**	**Unstandardised**
Fotokopieren	-0.82 (0.48)	19.84 (22.03)	1.07 (0.42)	90.29 (17.27)
Schlafwandeln	0.2 (0.79)	59.84 (34.99)	1.2 (0.34)	95.19 (12.49)

Figure 15

Density plots of verbs in Cluster 1 in both syntactic contexts



Cluster 2 (*n* = 6) follows a similar pattern but verbs in this group are not as highly acceptable in their non-split form (*M* = 0.08, *SD* = 0.19) as those in Cluster 1. Summaries of the average score per verb in this cluster are shown in Table 12 and Figure 16.

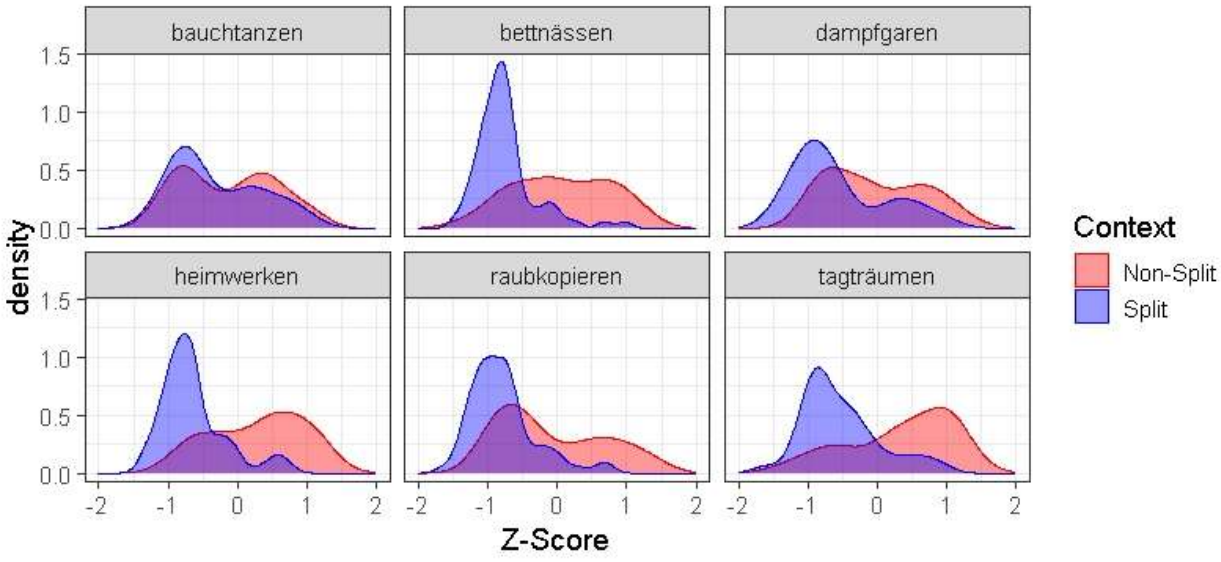
Table 12

Standardised and unstandardised means and standard deviations of verbs in Cluster 2. Standard deviations are given in parentheses

Verb	Split		Non-Split	
	Standardised	**Unstandardised**	**Standardised**	**Unstandardised**
Bauchtanzen	-0.29 (0.65)	36.09 (29.31)	-0.13 (0.68)	46.88 (32.17)
Bettnässen	-0.75 (0.44)	18.6 (22.87)	0.06 (0.71)	54.36 (31.66)
Dampfgaren	-0.55 (0.67)	29.51 (30.15)	-0.02 (0.69)	47.97 (31.6)
Heimwerken	-0.64 (0.47)	22.352 (23.36)	0.29 (0.67)	63.61 (28.54)
Raubkopieren	-0.79 (0.47)	21.14 (21.92)	-0.07 (0.76)	45.94 (34.33)
Tagträumen	-0.53 (0.56)	26.6 (26.58)	0.34 (0.76)	65.97 (31.86)

Figure 16

Density plots of the verbs in Cluster 2 in both syntactic contexts



Verbs in Cluster 3 (*n* = 5) show the reversal of this pattern by being acceptable in their split (*M* = 0.93, *SD* = 0.10) but not in the non-split form (*M* = -0.69, *SD* = 0.10). Their ratings are summarised in Table 13 and depicted in the density plots in Figure 17.

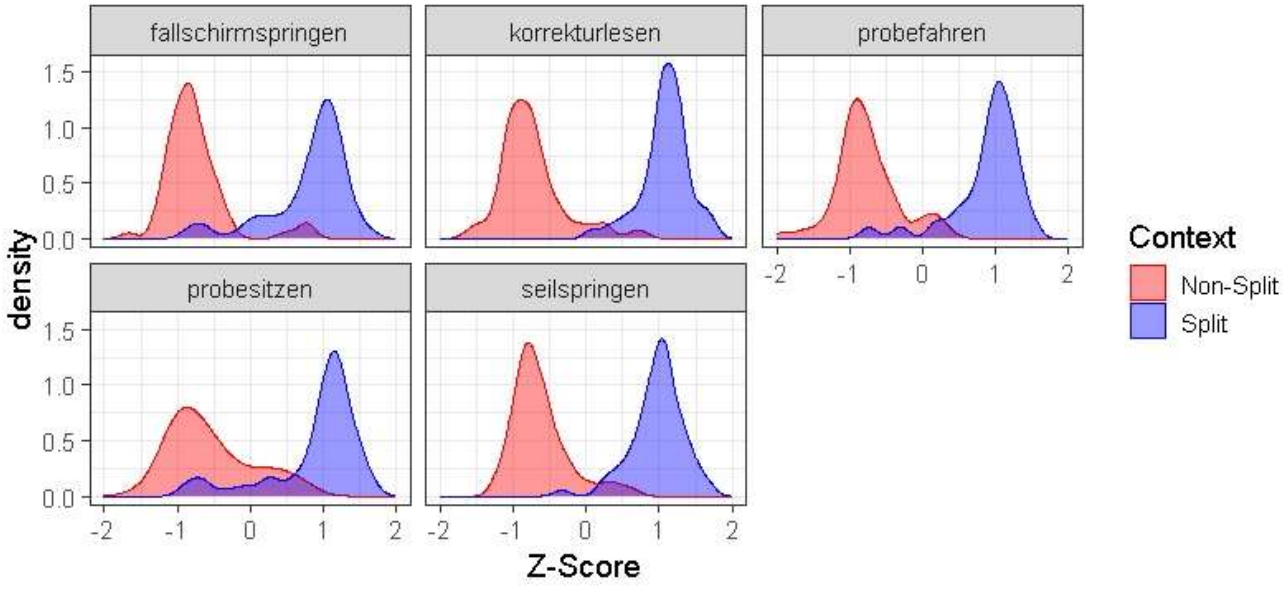
Table 13

Standardised and unstandardised means and standard deviations of verbs in Cluster 3. Standard deviations are given in parentheses

Verb	Split		Non-split	
	Standardised	**Unstandardised**	**Standardised**	**Unstandardised**
Fallschirmspringen	0.82 (0.54)	84.59 (22.98)	-0.76 (0.44)	18.28 (20.38)
Korrekturlesen	1.08 (0.33)	94.13 (9.97)	-0.75 (0.44)	19.41 (21.44)
Probefahren	0.89 (0.47)	87.93 (18.27)	-0.76 (0.44)	17.94 (20.67)
Probesitzen	0.89 (0.61)	83.24 (27.72)	-0.54 (0.57)	30.89 (27.46)
Seilspringen	0.98 (0.35)	90.78 (14.32)	-0.63 (0.39)	23.1 (21.71)

Figure 17

Density plots of verbs in Cluster 3 in both syntactic contexts



Cluster 4 (*n* = 5) is summarised in Table 14 and Figure 18 and contains verbs which are about as unacceptable in their non-split form (*M* = -0.68, *SD* = 0.14) as those in Cluster 3, but not as highly acceptable in their split form (*M* = 0.34, *SD* = 0.12) as those in Cluster 3.

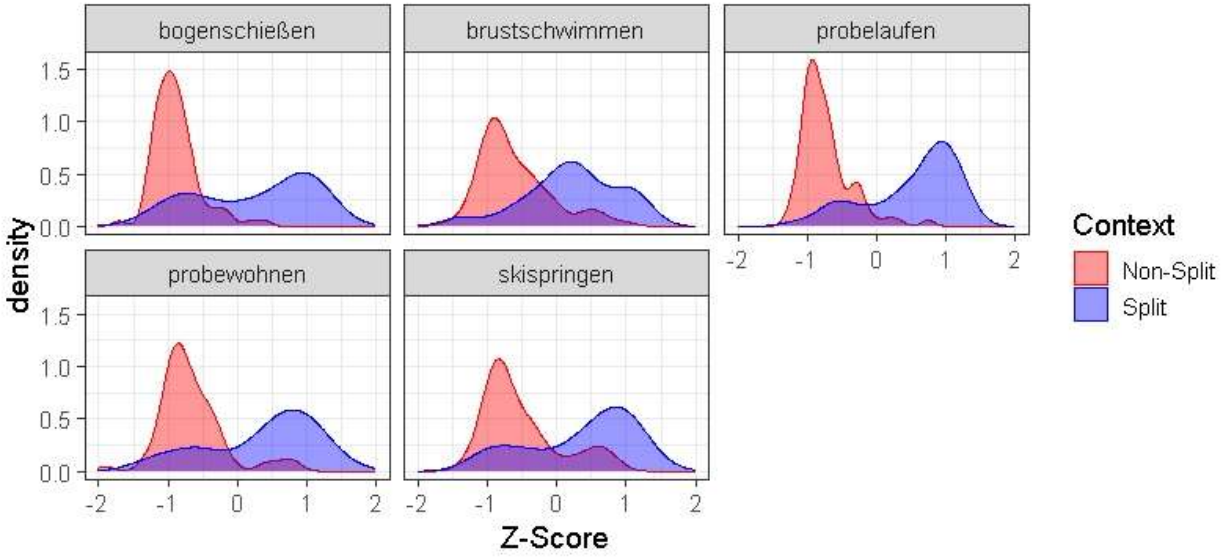
Table 14

Standardised and unstandardised means and standard deviations of verbs in Cluster 4. Standard deviations are given in parentheses

Verb	Split		Non-split	
	Standardised	**Unstandardised**	**Standardised**	**Unstandardised**
Bogenschießen	0.25 (0.83)	61.37 (35.52)	-0.88 (0.34)	14.12 (17.15)
Brustschwimmen	0.22 (0.69)	60.0 (32.31)	-0.6 (0.52)	26.03 (25.72)
Probelaufen	0.52 (0.65)	73.02 (29.16)	-0.73 (0.36)	19.34 (21.37)
Probewohnen	0.36 (0.79)	66.92 (33.12)	-0.66 (0.47)	22.46 (23.13)
Skispringen	0.36 (0.76)	67.37 (32.25)	-0.51 (0.56)	27.71 (26.86)

Figure 18

Density plots of the verbs in Cluster 4 in both syntactic contexts



Finally, for Clusters 5 and 6 the difference between the two syntactic configurations is not as prominent. The difference between those final two clusters lies in how highly both contexts were rated on average. Specifically, in Cluster 5 (n = 7), both contexts receive ratings around the centre of the scale (M_{split} = 0.13, SD_{split} = 0.25; $M_{non\text{-}split}$ = -0.04, $SD_{non\text{-}split}$ = 0.20). This can be seen in detail in Table 15 and the density plots in Figure 19.

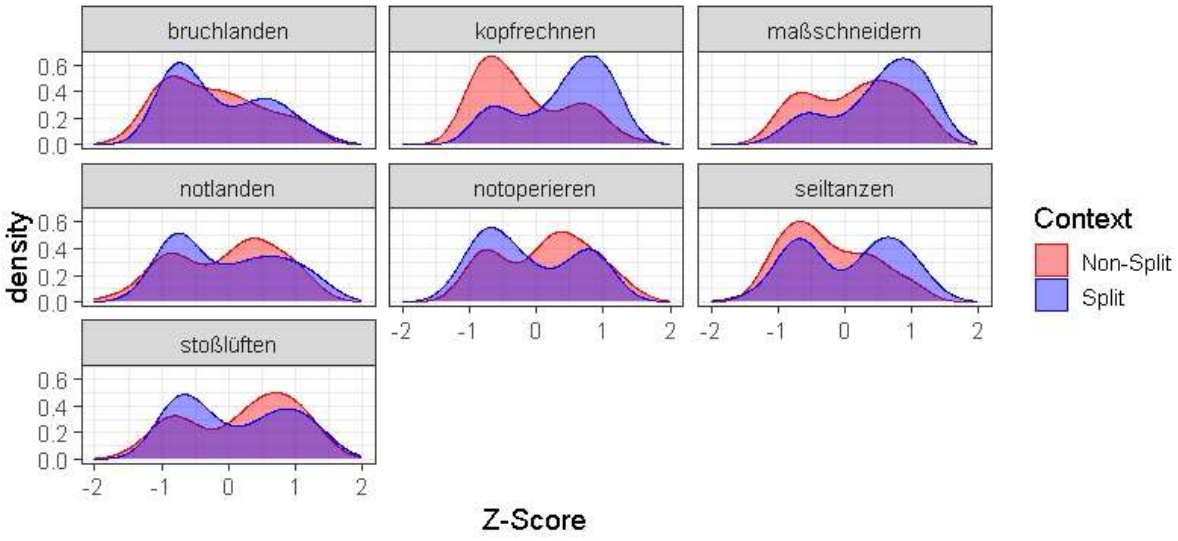
Table 15

Standardised and unstandardised means and standard deviations of verbs in Cluster 5. Standard deviations are given in parentheses

Verb	Split		Non-split	
	Standardised	**Unstandardised**	**Standardised**	**Unstandardised**
Bruchlanden	-0.14 (0.72)	43.029 (32.32)	-0.24 (0.73)	43.13 (32.86)
Kopfrechnen	0.41 (0.66)	68.85 (30.26)	-0.19 (0.69)	39.38 (31.99)
Maßschneidern	0.53 (0.67)	68.75 (31.17)	0.16 (0.69)	58.48 (30.56)
Notlanden	-0.007 (0.7)	47.75 (33.91)	-0.06 (0.81)	50.0 (34.67)
Notoperieren	-0.09 (0.73)	44.7 (32.17)	0.124 (0.7)	57.68 (30.69)
Seiltanzen	0.06 (0.76)	50.51 (33.19)	-0.25 (0.64)	42.99 (28.37)
Stoßlüften	0.11 (0.82)	52.03 (35.74)	0.2 (0.79)	59.18 (35.43)

Figure 19

Density plots of the verbs in Cluster 5 in both syntactic contexts



In contrast, in Cluster 6 (n = 5), both contexts are rejected by participants by receiving ratings lower than the average score (M_{split} = -0.48, SD = 0.17; $M_{non\text{-}split}$ = -0.54, SD = 0.19) which makes them the lowest rated verbs in both conditions. Table 16 and Figure 20 show detailed results per verb.

Table 16

Standardised and unstandardised means and standard deviations of verbs in Cluster 6. Standard deviations are given in parentheses

Verb	Split		Non-split	
	Standardised	**Unstandardised**	**Standardised**	**Unstandardised**
Bauchreden	-0.34 (0.68)	38.75 (31.45)	-0.32 (0.65)	35.01 (32.00)
Bergsteigen	-0.56 (0.64)	30.15 (28.45)	-0.48 (0.6)	29.35 (28.12)
Dauerparken	-0.6 (0.46)	24.22 (24.52)	-0.43 (0.65)	35.45 (28.79)
Ehebrechen	-0.65 (0.53)	22.81 (24.7)	-0.65 (0.56)	26.79 (24.31)
Eiskunstlaufen	-0.25 (0.67)	38.82 (29.19)	-0.79 (0.61)	21.48 (26.42)

Figure 20

Density plots of verbs in Cluster 6 in both syntactic contexts

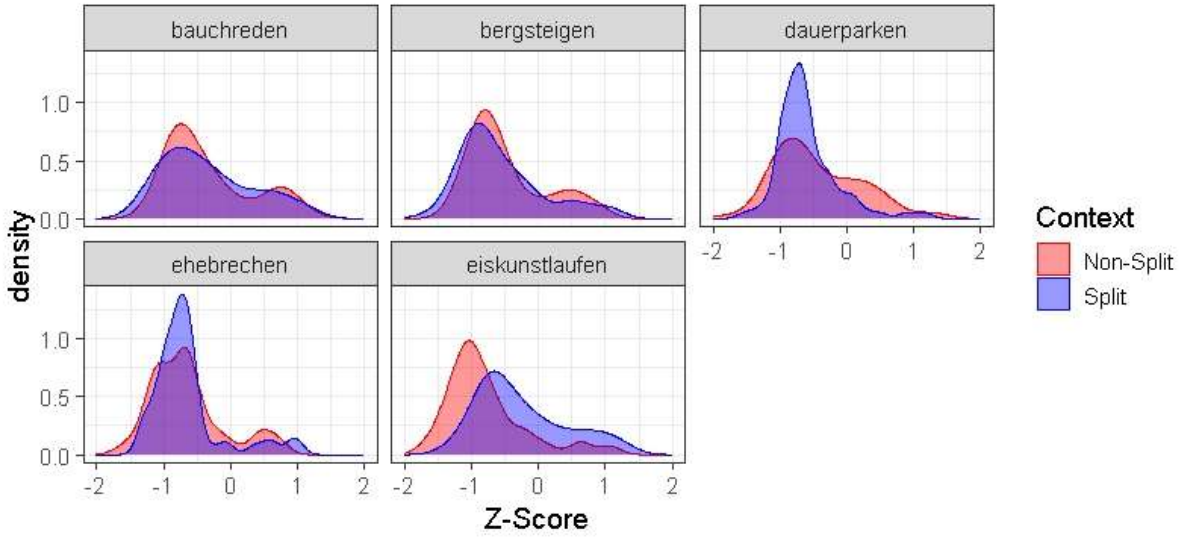
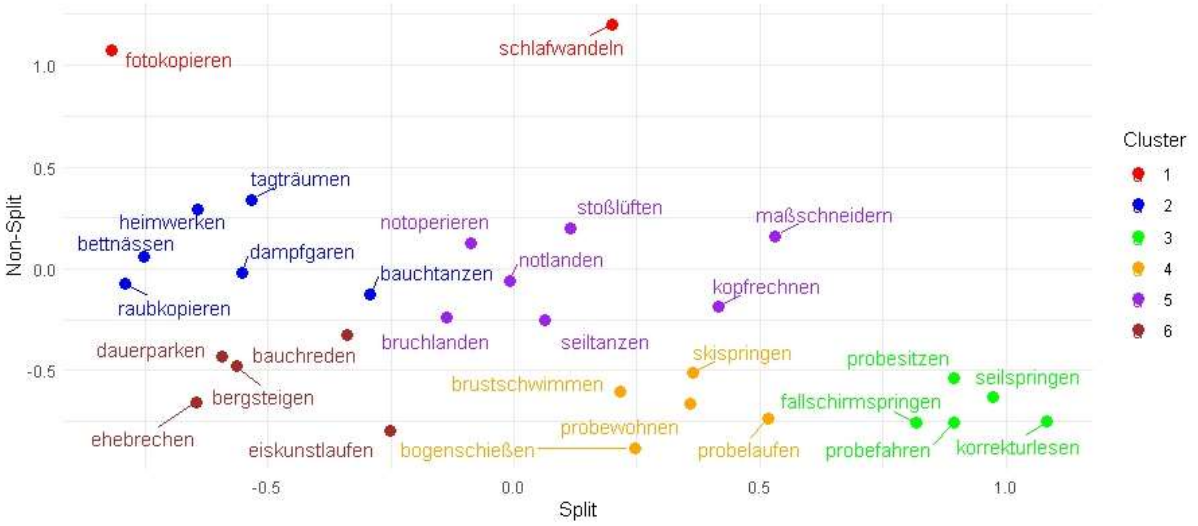


Figure 21 ilustrates where in the acceptability space each of the items is located. As also visible from the dendrogram in Figure 13, Figure 21 sheds light on the fact that even though the k-means clustering algorithm clusters verbs neighbouring each other in the space together, individual verbs may sometimes share more overlap with a member of a different cluster than with those in their own group. There is thus overlap in the rating space the clusters capture. This means that in the following discussion it should be kept in mind that even though the clustering results represent a useful aid in identifying underlying patterns, these should not be viewed as definitive categories inherent to the data.

Figure 21

Dot plot of mean rating per target verb in each context. Colours indicate which verbs were clustered together according to the k-means algorithm.

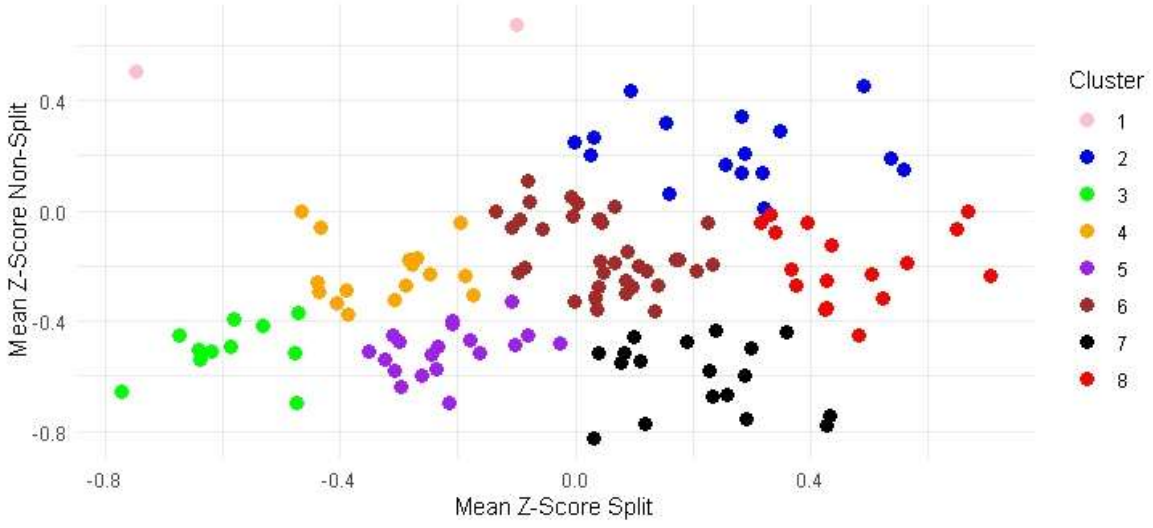


8.4.2 Participants

On participant level, the Silhouette method where k < 10 yielded nine as the optimal number of clusters. However, when constructing nine clusters, two of them only contained a single participant. In order to make visualisation more compact, those were collapsed into a single cluster. As can be seen in the plot in Figure 22, the two participants in this cluster were the ones to rate non-split contexts the most highly. Figure 23 visualises the final eight clusters as boxplots and Table 17 summarises the ratings per cluster.

Figure 22

Clustering of participants based on eight clusters. Dots represent individual participants' mean rating across target verbs in each context. Colours indicate which participants were clustered together according to the k-means algorithm



Aside from the already described Cluster 1 (*n* = 2), where non-split contexts received higher ratings (*M* = 0.59, *SD* = 0.12) than split contexts (*M* = –0.42, *SD* = 0.46), the highest overall ratings for both contexts were found in Cluster 2 (*n* = 36), the largest cluster. Participants in this group rated split contexts positively (*M* = 0.26, *SD* = 0.18) and non-split contexts similarly favourably (*M* = 0.23, *SD* = 0.12), thus not rejecting either context.

In contrast, the remaining clusters all show a clearer pattern of context preference or rejection. Clusters 3 to 5 consist of participants who rated both contexts below average. The most pronounced rejection is observed in Cluster 3 (*n* = 11), where ratings were lowest for both split (*M* = –0.59, *SD* = 0.09) and non-split (*M* = –0.50, *SD* = 0.10) contexts. Cluster 4 (*n* = 16) also shows below-average ratings, with split (*M* = –0.32, *SD* = 0.10) slightly lower than non-split (*M* = –0.22, *SD* = 0.11). Similarly, in Cluster 5 (*n* = 19), split contexts were rated *M* = –0.22 (*SD* = 0.09), while non-split ratings were even lower (*M* = –0.51, *SD* = 0.09).

A different trend emerges in Clusters 6 to 8, where split contexts were generally rated more favourably than non-split. In Cluster 6 (*n* = 18), both contexts received ratings close to the overall average, with split rated *M* = 0.05 (*SD* = 0.10) and non-split *M* = –0.15 (*SD* = 0.13). Cluster 7 (*n* = 17) exhibited the strongest rejection of non-split contexts (*M* = –0.60, *SD* = 0.13), while split ratings were noticeably more positive (*M* = 0.21, *SD* = 0.13). Finally, Cluster 8 (*n* = 16) displayed the highest split ratings across all clusters (*M* = 0.47, *SD* = 0.12), with non-split contexts again being rejected (*M* = –0.19, *SD* = 0.14).

To explore whether participants in the eight clusters differed in their demographics, Table 17 also includes the mean age per cluster. Here, it can be seen that Clusters 1 and 2, in which participants awarded items the highest rankings overall, had the youngest participants on average. In contrast, the oldest mean age can be found in Cluster 3.

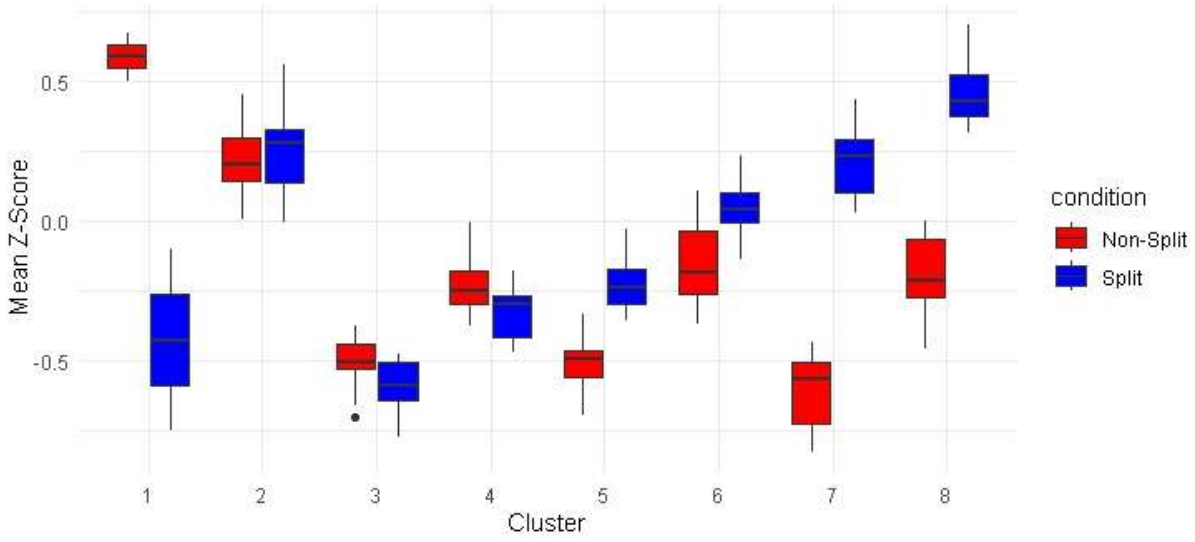
Table 17

Mean standardised scores in each condition per cluster of participants. Standard deviations are given in parentheses

Cluster	**1 (n = 2)**	**2 (n = 36)**	**3 (n = 11)**	**4 (n = 16)**	**5 (n = 19)**	**6 (n = 18)**	**7 (n = 17)**	**8 (n = 16)**
Split	-0.42 (0.46)	0.26 (0.18)	-0.59 (0.09)	-0.32 (0.10)	-0.22 (0.09)	0.05 (0.10)	0.21 (0.13)	0.47 (0.12)
Non-split	0.59 (0.12)	0.23 (0.12)	-0.50 (0.10)	-0.22 (0.11)	-0.51 (0.09)	-0.15 (0.13)	-0.60 (0.13)	-0.19 (0.14)
Mean age	27.5 (10.6)	37.5 (16.5)	53.9 (10.5)	41.3 (11.3)	45.9 (13.3)	40.3 (16.4)	39.3 (15)	42.4 (12.8)

Figure 23

Boxplots of ratings per cluster of participants in the two syntactic contexts.



9 Discussion

9.1 Research Question 1

The goal of the present study was to examine the properties and distribution of N + V verbs in German which have been hypothesised to be unable to appear in V2 position. A widely accepted, data driven description of which verbs are indeed non-V2 has been missing in the literature which is why RQ1 simply asked how target verbs are rated both in their split and non-split form and how those ratings compare to those of canonical complex verbs. Previous theoretical disagreements have raised the question of how real the phenomenon of non-V2 verbs is which made the present research necessary.

The study's results present a complex picture which means that RQ1 can be answered on multiple levels: as indicated by the results of Model 1 and the pairwise comparison, target items received significantly lower ratings than the two canonical grammatical patterns, non-split prefix verbs and split particle verbs. Now, these findings suggest that speakers do not judge target verbs to be fully acceptable in either of the two syntactic configurations, seemingly justifying their classification as *non-V2*. Importantly, this study is the first to experimentally compare the target verbs to regular prefix and particle verbs in a full-scale experiment with control conditions. Values reported here are thus more robust than the results of previous works and allow us to quantify the amount of deviation from the norm, which is especially important as it can be assumed that speakers judge items in relation to each other. Essentially, this means that if there are only target items in the experiment, low scores for some items will represent their deviation from the higher rated items within the same group of target verbs: by including canonical prefix and particle verbs, the study instead ensures that target verbs are assessed in comparison to clearly grammatical alternatives.

Despite the clear difference between target and non-target verbs, the terming of target verbs as *non-V2* cannot be straightforwardly assumed based on the present results. Here, I present four complicating factors for this terming. As a first point, thanks to the inclusion of control verbs in both syntactic contexts, it becomes evident that target verbs' results do not only significantly differ from control verbs

in the grammatical conditions but also in the respective ungrammatical conditions (non-split particle, split prefix). That is, they were not as fully unacceptable as clearly ungrammatical sentences in both verb categories. Instead, they received mixed ratings in both of those contexts. These results suggest participants may not treat the target verbs as grammatical violations. Given the clear use of opposing ends of the scale by speakers for the canonical verbs, ratings at the centre of the scale signal individual speakers' uncertainty about items' acceptability.

Secondly, Model 2 revealed a significant effect of pretest scores on the acceptability ratings of target verbs. This is important because theories of non-V2 verbs posit that these verbs should be acceptable in their finite form in non-V2 contexts. The pretest results already contradicted this and their predicative value for the main experiment further suggests that an alternative factor may underlie the low acceptability of these verbs in finite contexts overall. This aligns with Forche's (2020) claim that the relevant verbs do not appear in V2 not because of strict syntactic constraints, but because they tend to be rare in finite contexts more generally. These findings complicate a strict non-V2 classification, pointing instead to a broader issue in finite contexts.

One might hypothesise that items' low acceptability is driven by item frequency—i.e., that less frequent items are less familiar to speakers, causing uncertainty in how to use them which in turn results in the low ratings observed here. However, this hypothesis is not supported by the model results, which showed no significant effect of perceived frequency on ratings. One reason for the absence of such an effect may lie in the task design: participants judged how frequent the target item is in German but this may not reflect how often the word occurs as a *verb*. Since items were presented to participants in their infinitival form, they were often identical in surface form to a corresponding noun. As Forche (2020) notes, many of the verbs which have been described as *non-V2* are actually more frequent in their nominal or adjectival forms. This non-significant effect is also in line with Forche's own results who does not find a significant effect of corpus frequency on proportion of occurrence in V2 in his model. In combination with the pretest results, this means that it is possible that items do not standardly occur as finite verbs which then accounts for the results of both the pretest and the main experiment but would not necessarily be observable in the perceived frequency data. Whatever the reason for the ratings in both pretest and main experiment may be, their similarity suggests that they have to be regarded in unison which extends the issue faced by the verbs presented here to other finite contexts.

A third factor complicating target verbs' *non-V2* classification is item-level variation. The discussion so far has focused on general trends across all target verbs but a closer look reveals that individual items behave very heterogeneously. Some target verbs consistently receive low ratings, while others are judged more favourably – sometimes even approaching the acceptability of canonical particle or prefix verbs. This was illustrated in Figure 10 which depicts the mean z-score per target item in both contexts alongside the average z-scores of prefix and particle verbs.

The inclusion of random intercepts and by-context slopes for items improved the model fit for both linear mixed-effects models. This indicates that items differ both in their overall acceptability (intercepts), as well as in how strongly their ratings varied by syntactic context (slopes). In the subsequent cluster analysis, several distinct patterns emerged. This further supports the view that the target verbs do not behave as a homogeneous group. Consequently, the question of whether the verbs discussed in the literature qualify as *non-V2* cannot be answered uniformly for all of them but requires a detailed by-item analysis.

Fourthly, variation also exists between participants. The fit of both models was further improved by the inclusion of random slopes and intercepts for participants and the cluster analysis of the participant data revealed clusters on participant-level. This indicates that speakers vary in their judgements of verbs and that any given verb may display different rating patterns depending on the speaker.
The question of whether the verbs in this study can be classified as *non-V2* cannot be straightforwardly answered. As the observed item and participant variation had been expected from the literature review, RQ2 and RQ3 asked how this variation could be classified and whether the resulting classification could be linked to any item or participant level characteristics.

9.2 Research Question 2

The second research question aimed to investigate what patterns there are in the acceptability of target verbs and how those can be linked to verbs' morphological and semantic properties. In order to answer RQ2, the identified clusters on item level will be described one by one with the attempt of identifying potential factors driving this grouping.

9.2.1 Cluster 1
Acceptable in V2 as inseparable verbs – *fotokopieren*, *schlafwandeln*

The first cluster contained only two items which received highly favourable ratings in their split form which makes them the highest scoring verbs amongst all of the target verbs in either condition.

Fotokopieren 'photo.copy' is identified by Forche (2020) as a potential non-V2 verb, based on its assumed origin as a so-called pseudo-incorporated verb, as described by Wurzel (1998). In attempting to explain its nonconformity with this prediction, it is immediately noticeable that this verb stands out from all other tested verbs by one specific feature: as noted earlier, it is the only verb in the dataset with a deviant stress pattern—its primary stress falls on the penultimate syllable, rather than on the nominal element. As Eschenlohr (1999) explains, there is pressure in V2 contexts to split off the preverbal prefix when it bears the main word accent. Since this condition does not apply to *fotokopieren*, it avoids this structural pressure, thus explaining its greater acceptability in V2 contexts.

Schlafwandeln 'sleep.wander' = 'sleepwalk' received similarly high ratings, indicating its overall acceptability in V2 which contrasts with Eschenlohr (1999). Explaining why this verb does not exhibit a defective inflectional paradigm appears more challenging than in the case of *fotokopieren* 'photo.copy'. One possible account is found in Trommelen and Zonneveld (1986), who note that the Dutch equivalent *slaapwandelen* 'sleepwalk' can appear in V2. They attribute this to the fact that the base verb *wandelen* 'wander' has a weak inflectional paradigm. This aligns with Ahler's (2021) more recent *Head Inflection Hypothesis*, which posits that strongly inflected base verbs face two options: either the preceding element must be split from its base, or the entire compound is reanalysed as a simplex stem leading to a shift toward a weak inflectional paradigm. An example is the previously described *handhaben* 'hand.have' = 'handle', which is also acceptable in V2 as a whole (Ahlers, 2021). Its weak inflectional pattern is evident in its past tense form: the strong form *handhatte* is ungrammatical, while the weak form *handhabte* is acceptable, as shown in (23):

(23) a. **Er handhatte die Situation wie gewohnt.*
He handled the situation as usual.

b. *Er handhabte die Situation wie gewohnt.*
He handled the situation as usual.

Following Trommelen and Zonneveld (1986), weak inflection as in the case of *wandelen* and its German equivalent *wandern* would the make non-split appearance in V2 more likely.
Interestingly though, *schlafwandeln* was moderately acceptable (*M* = 0.2, *SD* = 0.79) even in its split form. This implies that grammatical representations of it may differ depending on the speaker and that some speakers instead reanalyse it as a regular particle verb.
Overall, although the accounts for why the two verbs in this cluster behave in this way differ, neither of these items can be classified as non-V2 and should not be included in future discussions of this phenomenon.

9.2.2 Cluster 2

Tendency towards V2 as inseparable verbs – *bettnässen*, *tagträumen*, *bauchtanzen*, *heimwerken*, *dampfgaren*, *raubkopieren*

The verbs in Cluster 2 were rejected in their split form but received mixed ratings in non-split contexts. This pattern suggests uncertainty in the usage of these verbs but indicates that overall, the non-split form is preferred.

The meaning of *heimwerken '*home.work' is intransparent in that it does not mean 'work at home' but rather describes a specific type of do-it-yourself crafting activity. The higher ratings in its non-split form therefore support a reanalysis theory as proposed by Fortmann (2007) according to which untransparent N + V combinations are more likely to appear in V2 as a whole.

The rejection of *raubkopieren '*theft.copy' = 'pirate' in V2 in its split form could be read as speakers forming an analogy to *fotokopieren* 'photo.copy'. In contrast, it is likely not as acceptable as the former in its non-split form due to its different stress pattern. Additionally, Forche (2020) argues that the *raub*-component may in fact be the stem of the verb *rauben* 'loot' and not the noun Raub 'theft' which would denote simultaneous looting and copying. A similar argument could hold for *dampfgaren* 'steam.cook' where *dampf-* could be the verbal stem of *dampfen* '(to) steam' rather than the noun *Dampf* 'steam'. Forche (2020) argues that such V + V compounds may be more acceptable in V2 than N + V compounds. However, such argumentation is highly speculative and it is therefore difficult to prove that speakers are indeed affected by such fine-grained morphological differences.

The final three verbs in this cluster – *bettnässen* 'bed.wet', *bauchtanzen* 'belly.dance', and *tagträumen* 'day.dream' – all exhibit semantically transparent structures: in *bettnässen*, the nominal and verbal components are in an argument relationship, while in *bauchtanzen* 'belly.dance' and *tagträumen* 'day.dream', the nominal elements function as modifiers of the base verb. According to Booij's (2005) account, argument relationships within complex predicates should allow split V2 appearances. In contrast, Fortmann (2007) argues that semantic transparency regardless of the specific relational type indicates a lack of lexicalisation, making such constructions less acceptable in V2 because they resist being reanalysed as lexical units. The fact that no clear difference in V2 acceptability emerges between verbs with argument relationships and those with modifier relationships lends more support to Fortmann's view.

9.2.3 Cluster 3

Acceptable in V2 as separable verbs – *fallschirmspringen*, *seilspringen*, *probesitzen*, *probefahren*, *korrekturlesen*

The third cluster of verbs – consisting of *fallschrimspringen* 'parachute.jump', *seilspringen* 'rope.jump', *probesitzen* 'trial.sit', *probefahren* 'trial.drive' and *korrekturlesen* 'correction.read' – is the one which behaves most homogenously. Ratings show a very clear overall bimodal distribution with high ratings for split and low ratings for non-split forms. This behaviour likens that of canonical particle verbs and verbs in this cluster should therefore not be classified as non-V2.

Semantically, all nominal elements in this cluster are modifiers of their base verbs. Their meaning is thus transparent. According to Fortmann (2007), transparency would hinder V2 occurrence altogether and according to Booij's (2007) account for Dutch, N + V verbs can only be split in V2 if the nominal element is in an argument relationship with the verb. Neither of these predictions are met by the present findings.

A better predictor for this cluster's behaviour appears to be provided by the *Head Inflection Hypothesis*: as was argued for *schlafwandeln* 'sleepwalk', weakly inflected base verbs may be more prone to reanalysis as a simplex stem (Ahlers 2020). Reversing this argument for the verbs presented in this cluster, it is striking that all four base verbs represented here are inflected strongly. In keeping with the argumentation presented for Cluster 1, this would mean that because verbs are inflected on their internal head, the preceding noun requires to be split which explains the low ratings in non-split cases.

9.2.4 Cluster 4

Tendency towards split occurrence – *probelaufen*, *probewohnen*, *skispringen*, *brustschwimmen*, *bogenschießen*

In Cluster 4, we find verbs whose structure shares some key features with those in Cluster 3. Specifically, *skispringen* 'ski.jump' shares the same base verb as the more acceptable *fallschirmspringen* 'parachute.jump' and *seilspringen* 'rope.jump' while *probelaufen* 'trial.walk' and *probewohnen* 'trial.live' share their nominal component with *probefahren* 'trial.drive' and *probesitzen* 'trial.sit' fro Cluster 3. Furthermore, as in the previous cluster, all nouns are in a modifier relationship with their base verb. This also includes the verbs *bogenschießen* 'arrow.shoot' and *brustschwimmen* 'breast.swim'.

In terms of ratings, these verbs' non-split forms are clearly rejected and their split version is preferred in all cases. However, their split ratings are not as high as those in Cluster 3. These ratings suggest that the factors leading to low scores in items' non-split form may be the same for Clusters 3 and 4 but that properties specific to Cluster 4 limit those verbs' acceptability when split.

The inclusion of all the *probe*- 'trial' verbs in either Cluster 3 or 4 is in line with a *reihenbildend* 'row building' account of complex verbs as purported by Eschenlohr (1999) according to which nominals that occur with verbs more frequently are more likely to be reanalysed as a regular separable particle which can be paired with a number ('row') of base verbs. The German reference grammar Duden lists verbs starting with *probe*- as examples of complex verbs which only appear in non-finite contexts (Dudenredaktion, 2016).

Furthermore, half of the verbs in Clusters 3 and 4 describe sportive activities, which is relevant for two reasons. First, Forche (2020) argues that sport-related verbs involve a high degree of agentivity, increasing the likelihood of these verbs developing finite forms. Second, he suggests that people who engage in sports may be more inclined to use such activities as verbs. More generally, this implies that N + V combinations denoting commonly performed actions may be more likely to appear in V2 due to pressure for communicative efficiency. This could explain why certain terms are more strongly embedded in individual speakers' grammars. However, this cannot be confirmed based on the present data as it would require information on participants' familiarity with the activities described by the verbs.

9.2.5 Cluster 5

Split and non-split at the centre of the scale – *stoßlüften*, *kopfrechnen*, *seiltanzen*, *notlanden*, *notoperieren*, *maßschneidern*, *bruchlanden*

In Cluster 5 the difference between split and non-split contexts is less pronounced as ratings for both contexts are clustered around the centre of the rating scale. Both the fact that both contexts received similar ratings and that neither context is clearly rejected by participants, indicates uncertainty in speakers about the "correct" use of these verbs in V2. Verbs in this cluster are *stoßlüften* 'hit.ventilate', *kopfrechnen* 'head.calculate', *seiltanzen* 'rope.dance', *notlanden* 'emergency.land', notoperieren 'emergency.operate', *maßschneidern* 'measurement.tailor' and *bruchlanden* 'break.land'. Ratings for these verbs often show a bimodal distribution in each syntactic context meaning that in both contexts some individual speakers rated them as acceptable while others rejected them more clearly.

It should be noted that some verbs in this cluster – most notably *kopfrechnen* 'head.calculate' (M = 0.41; SD = 0.66) and *maßschneidern* 'measurement.tailor' (M = 0.53; SD = 0.67) – received split ratings higher than the average of Cluster 4 (M = 0.31; SD = 0.12) which was argued to display a tendency towards acceptability in split contexts. Nevertheless, *kopfrechnen* and *maßschneidern* fell into Cluster 5 because their non-split forms were not rejected as clearly as those of verbs in Cluster 4. This suggests that speakers may have different sensitivities to the verbs' properties leading to different grammatical representations of these items.

9.2.6 Cluster 6

Genuine non-V2 verbs – *ehebrechen*, *bauchreden*, *eiskunstlaufen*, *bergsteigen*, *dauerparken*

Finally, a sixth of the verbs tested in the present study received negative z-scores both in their split as well as their non-split version. These make up their own cluster in the cluster analysis and it seems justified to classify them as non-V2 given the clear rejection of both contexts.

Among these are verbs commonly described as non-V2 in the literature, including *ehebrechen* 'marriage.break' = 'commit aduletery' and *bergsteigen* 'mountain.climb'. Identifying commonalities within this group proves challenging: the items do not share base verbs or nominal elements. Moreover, all of the semantic relationships identified within this group are also found in verbs in other clusters and therefore do not uniquely characterise any of the verbs in Cluster 6. For instance, one might expect that *bauchreden* 'belly.speak' – a verb in which a body part modifies the action of the base verb – would receive similarly high ratings to verbs such as *kopfrechnen* 'head.calculate' or *brustschwimmen* 'breast.swim' which were both more acceptable.

It seems that there is no uniform explanation for why the verbs in this last cluster refuse V2 position. Instead, answers more likely lie in individual properties of verbs. For example, as discussed previously, *ehebrechen* has been argued to be limited to biblical and infinitival contexts in modern language. The nominal component in *Eiskunst* 'ice.art' in *eiskunstlaufen* is bimorphemic which might result in an increased structural conflict.

Results in this cluster stress two important aspects: first of all, the fact that only five out of 30 verbs fall into this category, shows that the group of genuinely problematic verbs is smaller than what one might expect based on the theoretical literature. Secondly, the fact that no straightforward explanation for their behaviour surfaces points towards a more complex picture grounded in an interplay of multiple factors.

9.2.7 General Discussion

The item-level analysis shows that verbs form patterns based on their acceptability in split and non-split V2 contexts. This means that there is no single, homogeneous group of *non-V2* verbs identifiable in the dataset. RQ2 aimed to identify which factors drive this grouping. A key finding here is that the factors facilitating or hindering a verb's appearance in V2 depend on the syntactic context. This challenges earlier accounts such as Fortmann's (2007) semantic transparency hypothesis, which attempt to explain both split and non-split acceptability through a single approach. The present data suggest that acceptability ratings of both grammatical pathways must be independently accounted for. Consequently, results from studies that focus only on non-split forms, such as Ahlers (2021), are not as meaningful when compared to the current findings. Nevertheless, some of Ahlers' theoretical conclusions are of merit for the present data. In particular, his *Head Inflection Hypothesis* predicts that N + V verbs which are inflected on their internal head are required to be split – a prediction which aligns with the data from Cluster 3. However, the existence of verbs that show internal inflection but remain unacceptable in their split form such as *ehebrechen* 'marriage.break' = 'commit adultery' suggests that inflection alone cannot account for their behaviour. These cases may point towards additional processing constraints in these verbs' split versions.

Overall, the findings suggest that the syntactic behaviour of N + V verbs cannot be reduced to a single factor. Each cluster seems to be shaped by a range of interacting properties. In this regard, the most compatible previous account is Forche (2020), whose logistic regression analysis of corpus data also points to a complex explanation. Variables he identified which can also be hypothesised to play a role in the present results are stress pattern, degree of agentivity as well as whether the nominal component also pairs with other verbs. Secondly, Forche (2020) attests that such factors are not a sole predictor of items' V2 appearance but rather of their finite occurrence more generally.

Despite the possibility to identify some tentative grouping factors from the cluster analysis, it is also important to bear in mind that this analysis was only based on vectors of mean ratings per verb. This obscured the fact that some verbs' ratings have a higher spread than others which was presented in the results section and is also visible in the density plots displayed for each verb. A high spread of ratings indicates that speakers disagreed in their judgements. This is in line with the model comparison process which showed that participant factors are also relevant. Speakers thus display different preferences for which verbs can appear in V2 and whether this should be the case in their split or non-split form.

9.3 Research Question 3

RQ3 asked to what extent speakers vary in their judgements of N + V verbs in V2. To answer this question, several measures were undertaken. Firstly, Model 2 examining the ratings of target verbs only, found a significant effect of the demographic variable AGE. Older speakers rated the verbs in this study as lower than younger speakers in both contexts. Similarly, the
Intuitively, such a result may suggest a diachronic development of N + V verbs becoming more acceptable. However, one should be careful with such conclusions since the fact that this relationship existed for both contexts, does not provide evidence for the development of either split or non-split forms as the standard but rather points towards a property of the speakers responsible for these results such as older speakers' potential higher sensitivity to prescriptive violations or processing difficulties. Furthermore, the effect of age may also be mediated by level of education but this could not be tested for the present data.

Regional variety was not a variable in the model due to the large number of regions and small number of speakers for most regions. Given that complex verbs are a much-studied topic and that there is no notable publication suggesting dialectal microvariation in this regard, the absence of an effect of regional variety would be expected. Indeed, Vikner (2005) suggests the phenomenon of non-V2 verbs exists across the board in the OV West Germanic varieties including Standard German, Swiss German and Dutch.

While participants in this study cannot be characterised more concretely in terms of demographics, the emergence of groups based on participants' rating behaviour is evident from the cluster analysis. It confirmed that participants differed systematically in their preferences. A few participants clearly favoured non-split contexts (Cluster 1), while others consistently rejected both configurations (Clusters 3–5) or preferred the split form (Clusters 6 - 8). Interestingly, the largest group (Cluster 2) did not show a strong preference and rated both contexts positively. This suggests that acceptability is not only a matter of item-specific properties but also of how individual speakers process and evaluate these structures. The existence of clusters where both forms were rejected or both accepted further supports the idea that participants apply different strategies or constraints when judging N + V constructions in V2.

More concretely, as described, N + V verbs may present problems in both their split as well as their non-split form and the reasons for these problems differ for the two contexts. Individual data for participants suggests that the degree to which speakers are sensitive to these issues differs starkly. That is, it can be assumed that speakers who award higher ratings to verbs' split forms, are less impacted by a constraint which excludes non-arguments in verbs' argument spot as would be the case in verbs' split form, than by a constraint which prohibits a stressed element directly preceding the verb in V2. Conversely, participants who rated verbs in their non-split form more favourably, are not as sensitive to this issue. The group of participants who display this tendency appears to be smaller though than those whose follow the first rating pattern.

When jointly considering item and participant variation, it can be stated that the overall medium ratings for many of the target verbs generally stem from two sources: firstly, there is considerable variation in how speakers judged the items in both contexts: some find them fully acceptable while others arrive at the opposite judgement. Secondly, as evident from the speakers who tended to use the centre of the scale for their ratings of target verbs which indicates uncertainty in their judgements – speakers may perceive these constructions as odd sounding but not as fully ungrammatical.

10 Summary and Implications

The present study aimed to investigate the phenomenon of so-called non-V2 verbs in Standard German. As German is a V2 language, it is surprising that there are verbs which resist this parameter and there have been many accounts of these verbs' properties. A major concern with the theoretical literature was that there was no agreement about which verbs should even be classified as non-V2. This issue had not been resolved yet by experimental work as previous experiments only considered small sets of stimuli and did not compare target verbs to canonical verbs. This is what motivated the more thorough, theory-neutral examination that was presented in this thesis. The results of the study revealed that so-called non-V2 verbs do not form a homogeneous group with respect to their acceptability in V2 contexts. While some verbs were consistently rated low across participants and contexts, others received more favourable ratings sometimes approaching those of canonical particle or prefix verbs. Overall, seven out of 30 verbs received high ratings in one of the two contexts while many others clustered around the centre of the scale with only six out of 30 being rejected in both syntactic contexts more clearly. Importantly, there was also considerable item- and participant-level variation, with some participants rating target verbs much higher than others. More specific implications of the results range from 1) theoretical conclusions about complex verbs in German to 2) the nature of acceptability judgements and variability in participant data more broadly as well as to 3) more applied considerations in the domain of orthography.

Firstly, concerning the issue of complex N + V verbs, the present data provides a more nuanced perspective on the classification of so-called *non-V2 verbs*. Many authors such as Vikner (2005) and Fortmann (2007) have postulated non-V2 verbs to be a distinct type of complex verbs in German which exist as a third category alongside regular separable particle and inseparable prefix verbs. While ratings of verbs in this proposed category do significantly differ from the other two, I have argued that the present data does not support the existence of such a group as a stable grammatical category but rather points towards a more gradient phenomenon. This is because of the large amount of item and participant variation due to which no single theoretical account currently available captures the full range of variation observed. Overall, the results suggest that the status of these verbs as non-V2 items is often driven more by participant-specific variation than by fixed grammatical properties of the verbs themselves.

Bringing together all strands of the analysis, the final full account of the acceptability of target verbs examined in this study that I propose goes as follows: first, many of the so-called non-V2 verbs already display reduced acceptability in other finite contexts which the present study is the first to show experimentally. This reveals that prior research has arguably skipped a crucial step by focusing directly on verbs' V2 behaviour without first establishing their general acceptability in finite contexts. Next, when these verbs are placed in V2, they are subject to the structural ambiguity of having two potential grammatical pathways – either that of canonical particle verbs or that of prefix verbs. Since neither context received ratings on par with those of the grammatical control conditions, it can be assumed that both configurations impose additional constraints resulting in a reduction of acceptability in V2 on top of the already lower acceptability observed in other finite contexts. However, that fact that acceptability ratings differed starkly between both items and participants, suggests that there are factors which can mitigate the problems posed by these constraints. On item level, individual lexical properties such as stress patterns and inflectional class may play such a role while on participant level, speakers differ in how sensitive they are to the constraints imposed by each context. Finally, item and speaker variation may also interact. This might, for instance, be the case when speakers with higher

familiarity with specific items or the actions they denote may be hypothesised to award them higher ratings.

Extending from the field of complex verbs alone, the results present a strong case for the need for empirical data as well as the importance of variability in grammar judgements and individual variation. Findings align with a line of literature emphasising the importance of experimental data as a basis for linguistic theory building. For example, Featherston (2019) argues that experimental acceptability data is essential for capturing linguistic competence because introspective judgements by individual linguists are not sensitive enough. Findings here strongly support this claim as they show that in the case of so-called *non-V2 verbs*, judgments of individual theoretical authors were not sufficient to build a theory upon since they do not align with actual speaker data.

Moreover, results support the integration of individual variation into theoretical models. Specifically, they demonstrate that the variability in ratings observed across participants may not be random noise but rather reflects different underlying individual grammars. Similarly, Schoenmakers & van Hout (2024) find clear clusters of participants both regarding their ratings of norm violations as well as different configurations of scrambled and unscrambled constructions in Dutch. They argue that such patterns signal different underlying individual grammars. Such an account is especially relevant for the present data where items have two possible grammatical pathways at their disposal. Schoenmakers & van Hout (2024) further note that investigating the consistency and clustering of judgments is a useful tool for illuminating grammaticality patterns that are otherwise difficult to classify such as marginal grammaticality with ratings at the centre of the scale as also found in the present study. The cluster analysis presented here showed that when systematically analysed, seemingly messy data can become interpretable. The data presented in this thesis thus supports a perspective of a model of syntax that acknowledges variation both in analysis as well as in theory-building.

Lastly, since complex N + V verbs present an issue in German orthography, their study also has practical implications: the present experiment aimed to avoid a processing bias caused by speakers' expectations about noun capitalisation or words being spelled as one graphemic unit by presenting stimuli as audio recordings. Since native speakers of German have been found to be influenced by such spelling conventions (Lausch et al. 2021) in their lexical processing, conversely this also means that orthography should ideally align with speakers' representation of words. Concretely, for complex verbs this means that if a combination of a noun and a verb is interpreted as a single lexical unit akin to a regular prefix or particle verb, it would be spelled as one graphemic unit and the nominal part would – in contrast to usual German spelling – not be capitalised. Indeed, the German reference grammar Duden defines this rule as follows:

"If a noun appears as a verb affix, it depends on whether the substantival character is *perceived* as faded and the noun has lost its independence. If this is the case, it is written together." [1] **(Duden, n.d.; emphasis my own)**

It is interesting that the reference grammar describes the independence of the nominal element as *perceived*. This invariably triggers the question of whether this perception is shared between speakers.

[1] "*Tritt ein Substantiv als Verbzusatz auf, kommt es darauf an, ob der substantivische Charakter als verblasst empfunden wird und das Substantiv seine Eigenständigkeit verloren hat. Ist dies der Fall, wird zusammengeschrieben […]*"

While the existence of this orthographic rule creates the illusion that there is agreement about the status of these verbs, the use of word 'perceive' implies individual variation.

In some cases, German orthography does in fact allow for such variation to surface based on how individual speakers interpret the relationship between the verb's components. This can be best illustrated for the verb *staubsaugen* 'dust.suck' = 'vacuum clean' which is not usually seen as problematic, but which is generally accepted in two forms – as one graphemic unit as well as in the form *Staub saugen* 'dust suck' (Duden, n.d.) as presented in (24).

(24) a. *Ich staubsauge.*
I dust.suck

b. *Ich sauge Staub.*
I suck dust

'I vacuum clean.'

Given the previous definition, which version a speaker chooses would depend on whether they still perceive the noun *Staub* 'dust' as an independent argument of *saugen* 'suck' or whether the construction has become fully lexicalised. The data presented here suggest that it would be useful if similar orthographic solutions were available for some of the verbs in the present study particularly those with overall mixed ratings in both contexts. This is important given the relevance of both compound spelling and noun capitalisation in German orthography.

11 Limitations and Future Directions

As the study's main goal was to provide a broad picture of a range of different N + V constructions' acceptability, the selection of those tested constructions naturally lacks some depth. That is, target verbs were not selected using their internal semantic or morphological characteristics. The present study instead followed a bottom-up approach in which groups within the experimental target items were only identified based on the collected data. This is the opposite of the more common approach in experimental designs where items of interest are divided into groups first to then examine the effect of those grouping factors on the outcome variable. The order of steps taken in this research was deemed necessary in order to check the claims made in the literature. The resulting cluster analysis demonstrated a clear emergence of patterns within the verbs' ratings. This is in itself an important finding as it contradicts many previous theories which have postulated the existence of non-V2 verbs as a separate class of German complex verbs.

Nevertheless, the present findings remain inconclusive. While they show that verbs fall into distinct groups, they do not provide direct insight into the factors underlying this grouping as the experiment was not designed to allow for a robust statistical analysis of internal properties such as the verbs' semantic or morphosyntactic characteristics which is why all conclusions drawn here in this regard are still speculative. Although the results point toward accounts that emphasise individual variation, this variation may be driven by different, interacting factors. A more fine-grained analysis of these factors would be necessary to draw definitive conclusions. Future research could build on the patterns tentatively identified here to develop experimental conditions. Another potential approach would be to use the random effects extracted from the present model as outcome variables in subsequent analyses. Finally, in light of the low ratings observed already in the pretest, it seems reasonable to pivot

future research more towards the broader question of non-finiteness and the development of verbal inflectional paradigms.

For this reason, a second line of potential further research would be verbs' behaviour in other syntactic contexts. Results here only showed whether the N element could be split from the verb in standard present tense V2 contexts. While this answered the question of whether the concerned verbs can appear in V2 or not, the experiment did not include other contexts in which these verbs would be split. In order to gain a full understanding of how close the union between N and V is in those verbs, it would be interesting to examine further syntactic contexts. Such conditions of interest would be fronted particles as well as affixation.

Firstly, under an account such as that brought forward by Ahlers (2021), split nominal elements in N + V verbs are unacceptable because they do not align with the parser's processing expectations. As argued by Song (2021), this would predict that these verbs are separable in principle but not in V2. Song (2021) therefore predicts fronting to be more acceptable than the same verb split in V2. This exact pattern is reported by Terboven (2020) who reports higher ratings for the verb *ehebrechen* when fronted as compared to being split in V2. Experimentally testing this hypothesis in more depth would shed light on the processing-based account of the non-separability.

Secondly, verbs' so-called morphological separability is relevant because for regular prefix and particle verbs there are two distinct pathways analogous to their behaviour in V2. Particularly, separable particle verbs are split by the past tense morpheme and the infinitival marker *zu 'to'*. Clear preferences for one or the other in the case of target verbs might therefore be an indicator of the path the verb would take when used in V2. This is especially interesting for the verbs which received mixed ratings in both contexts. Methodologically, a cluster analysis like the one presented seems suitable for this line of research and would also be insightful if more properties of the items were examined.

Another field worthy of investigation is how the N + V verbs behave cross-linguistically. Complex verbs are an area which have received considerable attention in the study of microvariation of Germanic languages. Descriptive theoretical data by Booij (2005), Koopman (1995) and others demonstrate the same phenomenon to exist in Dutch but, again, there is currently no broad empirical description which clarifies the extent of this claim. The current results suggest that the verbs' ratings are driven much more by individual differences and different pathways to grammaticality. Such an approach would be unlikely to predict cross-linguistic grammatical differences between German and Dutch in this regard, but this would have to be tested empirically.

In conclusion, the goal of this thesis was to empirically verify the claim that there are German verbs that cannot appear in V2. Results show that for N + V verbs, such a claim cannot be made across the board. Instead, both item and participant variation lead to an overall medium score for target verbs. Future work should examine the factors underlying this variation in more detail and extend research to more syntactic contexts and languages.

List of Abbreviations

3	third person
PART	particle
PTCP	participle
PREF	prefix
REFL	reflexive
SG	singular

List of References

Ahlers, T. (2021). Gelobpreist oder lobgepriesen? Sind Pseudokomposita eigentlich Verbkomposita? Korpus- und Experimentdaten zur Kopfflexionsthese bei immobilen Verben des Deutschen. *Wiener Linguistische Gazette, 89*, 9–53. Retrieved from https://wlg.univie.ac.at/archiv/festschrift-892021/

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*(4), 584–589. https://doi.org/10.3758/BF03193889

Dudenredaktion. (n.d.). Fotokopieren. In *Duden online*. Retrieved January 28, 2025, from https://www.duden.de/rechtschreibung/fotokopieren

Dudenredaktion. (n.d.). Getrennt- und Zusammenschreibung von Verbindungen aus Substantiv und Verb. In *Duden online*. Retrieved May 15, 2025, from https://www.duden.de/sprachwissen/sprachratgeber/Getrennt-und-Zusammenschreibung-von-Verbindungen-aus-Substantiv-und-Verb

Dudenredaktion. (n.d.). Staub saugen. In *Duden online*. Retrieved January 28, 2025, from https://www.duden.de/rechtschreibung/staub_saugen

Dudenredaktion. (n.d.). Staubsaugen. In *Duden online*. Retrieved January 28, 2025, from https://www.duden.de/rechtschreibung/staubsaugen

Eisenberg, P. (2004). *Grundriss der deutschen Grammatik* (Band 1: Das Wort, 2., überarbeitete und aktualisierte Auflage). Metzler.

Eschenlohr, S. (1999). *Vom Nomen zum Verb: Konversion, Präfigierung und Rückbildung im Deutschen* (Vol. 3). Georg Olms Verlag.

Featherston, S. (2020). Can we build a grammar on the basis of judgments. In L. Juge & R. A. Laury (Eds.), *Linguistic intuitions: Evidence and method* (pp. 165–188). Oxford University Press. https://doi.org/10.1093/oso/9780198840558.003.0010

Forche, C. R. (2020). *NonV2-Verben im Deutschen: Theoretische Überlegungen und empirische Untersuchungen zu einem morphosyntaktischen Problemfall (den es vielleicht gar nicht gibt)*. Springer. https://doi.org/10.1007/978-3-662-61926-1

Fortmann, C. (2015). Verbal pseudo-compounds in German. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word formation: An international handbook of the languages of Europe* (Vol. 1, pp. 594–610). De Gruyter. https://doi.org/10.1515/9783110246254-036

Freywald, U., & Simon, H. J. (2007). Wenn die Wortbildung die Syntax stört: Über Verben, die nicht in V2 stehen können. In C. Fleischer & H. J. Simon (Eds.), *Verbale Wortbildung im Spannungsfeld zwischen Wortsemantik, Syntax und Rechtschreibung* (pp. 181–194). Walter de Gruyter.

Gerdts, D. B. (2017). Incorporation. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 84–100). Wiley-Blackwell.

Grimm, J. (1878). *Werke: Deutsche Grammatik* (Vol. 2). Dietrich.

Hüning, M. (2018). Foreign word-formation in Construction Morphology: verbs in-ieren in German. In G. Booij (Ed.), *The construction of words: Advances in Construction Morphology* (pp. 341–371). Springer. https://doi.org/10.1007/978-3-319-74394-3_13

Leibniz-Institut für Deutsche Sprache. (2024). *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2024-I* (Release: 13.03.2024). Retrieved from https://www.ids-mannheim.de/dereko

Koopman, H. (1995). On verbs that fail to undergo V-second. *Linguistic Inquiry, 26*(1), 137–163.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Labusch, M., Kotz, S. A., & Perea, M. (2022). The impact of capitalized German words on lexical access. *Psychological Research*, 1–12. https://doi.org/10.1007/s00426-021-01540-3

Lenth, R. (2023). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.8.6). Retrieved from https://CRAN.R-project.org/package=emmeans

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company. https://doi.org/10.1075/z.195

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2025). *cluster: Cluster analysis basics and extensions* (Version 2.1.8.1). Retrieved from https://CRAN.R-project.org/package=cluster

McIntyre, A. (2002). Idiosyncrasy in particle verbs. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-particle explorations* (pp. 95–118). Mouton de Gruyter. https://doi.org/10.1515/9783110902341.95

McIntyre, A. (2007). Particle verbs and argument structure. *Language and Linguistics Compass, 1*(4), 350–367. https://doi.org/10.1111/j.1749-818X.2007.00013.x

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, pp. 435–449). De Gruyter. https://doi.org/10.1515/9783110246254-025

Müller, S. (2002). *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German*. CSLI Publications.

Näser, W. (n.d.). Heinz Erhardt (1909–1979) zum Gedächtnis: Sinniges und Hintersinniges. [Webpage]. Retrieved January 28, 2025, from https://wolfgang-naeser-marburg.lima-city.de/htm/erhardt.htm

Rat für deutsche Rechtschreibung. (2024). *Amtliches Regelwerk der deutschen Rechtschreibung: Regeln und Wörterverzeichnis*. Leibniz-Institut für Deutsche Sprache.

Schäfer, R., & Sayatz, U. (2024). Between syntax and morphology: German noun+verb units. *Glossa: A Journal of General Linguistics, 9*(1). https://doi.org/10.16995/glossa.10069

Schoenmakers, G. J. (2023). Linguistic judgments in 3D: The aesthetic quality, linguistic acceptability, and surface probability of stigmatized and non-stigmatized variation. *Linguistics, 61*(3), 779–824. https://doi.org/10.1515/ling-2021-0179

Schoenmakers, G. J., & van Hout, R. (2024). Consistency and variability in acceptability judgments from naive native speakers. *Nederlandse Taalkunde, 29*(1), 49–75. https://doi.org/10.5117/NEDTAA2024.1.005.SCHO

Song, C. (2019). *On the formal flexibility of syntactic categories*. Doctoral dissertation, University of Cambridge.

Terboven, F. (2021). *Wort oder Wortteil? Empirische Untersuchung zum Status der Erstglieder in Nomen-Verb-Verbindungen*. Doctoral dissertation, Bergische Universität Wuppertal. https://doi.org/10.25926/kq21-5j83

Vikner, S. (2005). Immobile complex verbs in Germanic. *Journal of Comparative Germanic Linguistics, 8*, 83–115. https://doi.org/10.1007/s10828-004-0726-9

Trommelen, M., & Zonneveld, W. (1986). Dutch morphology: Evidence for the right-hand head rule. *Linguistic Inquiry, 17*(1), 147–169.

Zwischenstaatliche Kommission für deutsche Rechtschreibung. (2004). *Deutsche Rechtschreibung: Regeln und Wörterverzeichnis. Amtliche Regelung. Überarbeitete Fassung 2004*. Institut für Deutsche Sprache.

Appendix A

Table A1

List of verbs in the pretest including their total corpus frequency and their standardised and unstandardised pretest ratings. Standard deviations are given in parentheses

Verb	Gloss	Total corpus frequency	Pretest	Pretest Score
			Standardised	Unstandardised
abendessen	evening.eat	444	-0.99 (1.21)	50.21 (38.74)
bauchpinseln	belly.brush	267	-1.04 (0.86)	35.25 (31.78)
bauchreden	belly.speak	76	0.09 (0.78)	73.57 (28.50)
bauchtanzen	belly.dance	89	0.06 (0.76)	74.10 (29.34)
bausparen	building.save	120	-0.99 (1.01)	44.90 (35.05)
bergsteigen	mountain.climb	1048	0.17 (0.60)	76.86 (25.83)
bergwandern	mountain.hike	195	-0.43 (0.73)	52.44 (25.45)
bettnässen	bed.wet	46	0.19 (0.77)	80.05 (26.28)
bogenschießen	arrow.shoot	83	0.30 (0.70)	77.69 (23.24)
bruchlanden	break.land	112	-0.21 (0.93)	63.67 (33.43)
brustschwimmen	breast.swim	76	0.10 (0.75)	81.32 (26.95)
dampfgaren	steam.cook	135	-0.80 (1.45)	63.21 (30.14)
dauerlaufen	duration.walk	49	-0.66 (0.75)	44.38 (31.62)
dauerparken	duration.park	91	0.07 (0.80)	66.81 (33.01)
dienstverpflichten	service.obligate	160	-1.40 (0.54)	21.31 (22.83)
ehebrechen	marriage.break	599	-0.56 (0.76)	60.79 (36.43)
eisklettern	ice.climb	52	-0.51 (0.87)	58.33 (30.19)
eiskunstlaufen	ice.art.walk	112	-0.05 (0.87)	64.50 (31.97)
fallschirmspringen	parachute.jump	74	0.08 (0.79)	69.44 (32.09)
fotokopieren	photo.copy	365	0.16 (0.84)	71.94 (30.36)
gewichtheben	weight.lift	387	-1.05 (1.36)	48.74 (37.22)
haftpflichtversichern	liability.insure	72	-1.59 (1.09)	31.90 (23.98)
handarbeiten	hand.work	1339	-0.26 (0.83)	58.56 (28.60)
heimwerken	home.work	289	-0.45 (0.94)	63.58 (38.29)
irreführen	delusion.guide	1635	-0.67 (0.78)	53.14 (34.76)
irreleiten	delusion.lead	264	-0.89 (1.05)	50.76 (31.79)
kopfrechnen	head.calculate	326	0.34 (0.49)	84.76 (17.87)
kopfschütteln	head.shake	136	-0.61 (1.04)	57.35 (30.06)
korrekturlesen	correction.read	210	0.34 (0.73)	87.74 (25.15)
lobsingen	praise.sing	125	-0.42 (0.80)	51.38 (31.27)
luxussanieren	luxury.sanify	122	-1.04 (0.47)	31.94 (25.10)
manndecken	man.cover	175	-2.34 (0.98)	19.00 (22.35)
maßschneidern	measurement.sew	1812	0.12 (0.75)	70.56 (31.28)
notbremsen	emergency.brake	438	-0.29 (0.77)	57.13 (30.08)
notlanden	emergency.land	12960	0.22 (0.76)	83.79 (26.27)
notoperieren	emergency.operate	124	0.22 (0.71)	79.62 (27.69)
notschlachten	emergency.butcher	495	-0.86 (1.12)	37.44 (32.77)
notverkaufen	emergency.sell	82	-0.70 (0.55)	42.44 (23.86)
notversorgen	emergency.provide	113	-0.71 (0.96)	56.05 (38.90)
notwassern	emergency.water	600	-1.33 (0.68)	25.19 (22.05)
probefahren	trial.drive	885	0.14 (0.75)	81.16 (24.49)
probelaufen	trial.walk	80	-0.02 (0.59)	77.68 (24.34)
probesitzen	trial.sit	1509	0.19 (0.45)	82.63 (21.89)

probewohnen	trial.live	243	-0.68 (0.95)	59.95 (36.65)
rätselraten	mystery.guess	54	-0.73 (0.95)	43.69 (29.12)
raubkopieren	theft.copy	88	-0.23 (0.84)	70.16 (30.97)
sackhüpfen	sack.jump	83	-0.12 (0.55)	57.00 (27.50)
sandstrahlen	sand.stream	401	-0.69 (0.89)	52.10 (33.62)
schaulaufen	show.walk	221	-0.75 (0.78)	51.14 (32.12)
schlafwandeln	sleep.wander	596	0.44 (0.58)	85.43 (23.14)
schockfrosten	shock.freeze	140	-0.80 (0.68)	47.70 (32.57)
schutzimpfen	protection.vaccinate	50	-0.78 (0.98)	56.50 (33.16)
segelfliegen	sail.fly	238	-0.79 (0.68)	41.00 (28.57)
seilspringen	rope.jump	140	0.24 (0.50)	83.89 (23.18)
seiltanzen	rope.dance	104	0.28 (0.67)	83.89 (29.79)
skispringen	ski.jump	158	0.24 (0.61)	73.44 (26.06)
stoßlüften	hit.ventilate	123	0.37 (0.53)	90.53 (12.99)
tagträumen	day.dream	60	0.13 (0.67)	82.68 (27.28)
wahlkämpfen	election.fight	450	-1.63 (0.84)	26.67 (26.22)
wärmedämmen	warmth.dam	157	-0.80 (0.81)	49.10 (29.98)

Appendix B

Table B1

List of experimental stimuli in the main experiment (excluding filler items)

Verb type	Syntactic context		Subject	Verb	Adverb/Object	Particle
target	non-split	Die	Frau	bauchtanzt	manchmal	
target	split	Die	Frau	tanzt	manchmal	bauch
		‘The woman sometimes belly dances.‘				
particle	non-split	Die	Frau	einkauft	manchmal	
particle	split	Die	Frau	kauft	manchmal	ein
		‘The woman sometimes goes shopping.‘				
prefix	non-split	Die	Frau	verkauft	das Buch	
prefix	split	Die	Frau	kauft	das Buch	ver
		‘The woman sells the book.‘				
target	non-split	Der	Anwohner	dauerparkt	noch	
target	split	Der	Anwohner	parkt	noch	dauer
		‘The resident still parks permanently.’				
particle	non-split	Der	Anwohner	umbaut	noch	
particle	split	Der	Anwohner	baut	noch	um
		‘The resident still rebuilds.‘				
prefix	non-split	Der	Anwohner	bebaut	das Grundstück	
prefix	split	Der	Anwohner	baut	das Grundstück	be
		‘The resident builds on the property site.‘				
target	non-split	Der	Akrobat	seiltanzt	oft	
target	split	Der	Akrobat	tanzt	oft	seil
		‘The acrobat often rope dances.‘				
particle	non-split	Der	Akrobat	nachdenkt	oft	
particle	split	Der	Akrobat	denkt	oft	nach
		‘The acrobat often ponders.‘				
prefix	non-split	Der	Akrobat	bedenkt	die Choreographie	
prefix	split	Der	Akrobat	denkt	die Choreographie	be

‘The acrobat considers the choreography.‘

target	non-split	Der	Mann	ehebricht	manchmal	
target	split	Der	Mann	bricht	manchmal	ehe

‘The man sometimes commits adultery.‘

particle	non-split	Der	Mann	abstimmt	manchmal	
particle	split	Der	Mann	stimmt	manchmal	ab

‘The man sometimes votes.‘

prefix	non-split	Der	Mann	verstimmt	die Gitarre	
prefix	split	Der	Mann	stimmt	die Gitarre	ver

‘The man puts the guitar out of tune.‘

target	non-split	Die	Designerin	maßschneidert	oft	
target	split	Die	Designerin	schneidert	oft	maß

‘The designer often tailors.‘

particle	non-split	Die	Designerin	einwilligt	oft	
particle	split	Die	Designerin	willigt	oft	ein

‘The deisgner often agrees.‘

prefix	non-split	Die	Designerin	bewilligt	den Entwurf	
prefix	split	Die	Designerin	willigt	den Entwurf	be

‘The designer agrees to the draft.’

target	non-split	Der	Zweijährige	bettnässt	oft	
target	split	Der	Zweijährige	nässt	oft	bett

‘The two-year old often bed wets.’

particle	non-split	Der	Zweijährige	aufsteht	oft	
particle	split	Der	Zweijährige	steht	oft	auf

‘The two-year old often gets up.

prefix	non-split	Der	Zweijährige	versteht	das Spiel	
prefix	split	Der	Zweijährige	steht	das Spiel	ver

‘The two-year old understands the game.’

target	non-split	Der	Großvater	tagträumt	noch	
target	split	Der	Großvater	träumt	noch	tag

‘The grandfather still daydreams.‘

particle	non-split	Der	Großvater	ansteht	noch	
particle	split	Der	Großvater	steht	noch	an

‘The grandfather still queues.‘

prefix	non-split	Der	Großvater	besteht	den Sehtest	
prefix	split	Der	Großvater	steht	den Sehtest	be

‘The grandfather passest he eye examination.‘

target	non-split	Der	Schütze	bogenschießt	oft	
target	split	Der	Schütze	schießt	oft	bogen

‘The shooter often arrow shoots.‘

particle	non-split	Der	Schütze	aufgibt	oft	
particle	split	Der	Schütze	gibt	oft	auf

‘The shooter often gives up.‘

prefix	non-split	Der	Schütze	vergibt	die Chance	
prefix	split	Der	Schütze	gibt	die Chance	ver

‘The shooter gives up on the chance.‘

target	non-split	Der	Athlet	brustschwimmt	wieder	
target	split	Der	Athlet	schwimmt	wieder	brust

‘The athelete breast swims

particle	non-split	Der	Athlet	mitgeht	wieder	
particle	split	Der	Athlet	geht	wieder	mit

‘The athelete joins again.‘

prefix	non-split	Der	Athlet	begeht	den Fehler	
prefix	split	Der	Athlet	geht	den Fehler	be

‘The athelete makes the mistake.‘

target	non-split	Der	Wanderer	bergsteigt	wieder	
target	split	Der	Wanderer	steigt	wieder	berg

‘The hiker mountaineers.‘

particle	non-split	Der	Wanderer	zurückkommt	wieder	
particle	split	Der	Wanderer	kommt	wieder	zurück

‘The hiker returns again.‘

prefix	non-split	Der	Wanderer	bekommt	die Medaille	
prefix	split	Der	Wanderer	kommt	die Medaille	be

‘The hiker receives the medal.‘

target	non-split	Die	Abenteurerin	fallschirmspringt	noch	
target	split	Die	Abenteurerin	springt	noch	fallschirm

‘The adventurer still parachutes.‘

particle	non-split	Die	Abenteurerin	zurechtkommt	noch	
particle	split	Die	Abenteurerin	kommt	noch	zurecht

‘The adventurer still gets along.‘

prefix	non-split	Die	Abenteurerin	entkommt	der Gefahr	
prefix	split	Die	Abenteurerin	kommt	der Gefahr	ent

‘The adventurer escapes the danger.‘

target	non-split	Die	Nachbarin	schlafwandelt	manchmal	
target	split	Die	Nachbarin	wandelt	manchmal	schlaf

‘The neighbour sometimes sleepwalks.‘

particle	non-split	Die	Nachbarin	abschließt	manchmal	
particle	split	Die	Nachbarin	schließt	manchmal	ab

‘The neighbour sometimes locks.‘

prefix	non-split	Die	Nachbarin	verschließt	die Tür	
prefix	split	Die	Nachbarin	schließt	die Tür	ver

‘The neighbour locks the door.‘

target	non-split	Der	Student	probewohnt	wieder	
target	split	Der	Student	wohnt	wieder	probe
		‘The student				
particle	non-split	Der	Student	ausgeht	wieder	
particle	split	Der	Student	geht	wieder	aus

‘The student goes out again.‘

prefix	non-split	Der	Student	entgeht	der Nachprüfung	
prefix	split	Der	Student	geht	der Nachprüfung	ent

‘The student escapes the resit.‘

target	non-split	Der	Urlauber	skispringt	wieder	
target	split	Die	Urlauberin	springt	wieder	ski

‘The tourist does ski jumping again.‘

particle	non-split	Die	Urlauberin	aufwacht	wieder	
particle	split	Die	Urlauberin	wacht	wieder	auf

‘The tourist wakes up again.‘

prefix	non-split	Die	Urlauberin	bewacht	das Gepäck	
prefix	split	Die	Urlauberin	wacht	das Gepäck	be

‘The tourist watches the luggage.‘

target	non-split	Die	Joggerin	probeläuft	noch	
target	split	Die	Joggerin	läuft	noch	probe

‘The jogger

particle	non-split	Die	Joggerin	durchhält	noch	
particle	split	Die	Joggerin	hält	noch	durch

‘The jogger still hangs on.‘

prefix	non-split	Die	Joggerin	behält	die Wasserflasche	
prefix	split	Die	Joggerin	hält	die Wasserflasche	be

‘The jogger keeps the water bottle.‘

target	non-split	Die	Rentnerin	heimwerkt	wieder	
target	split	Die	Rentnerin	werkt	wieder	heim

‘The pensioner crafts again.‘

particle	non-split	Die	Rentnerin	umzieht	wieder	
particle	split	Die	Rentnerin	Zieht	wieder	um

‘The pensioner moves again.‘

prefix	non-split	Die	Rentnerin	bezieht	das Bett	
prefix	split	Die	Rentnerin	zieht	das Bett	be

‘The pensioner makes the bed.‘

target	non-split	Die	Mutter	stoßlüftet	wieder	
target	split	Die	Mutter	lüftet	wieder	stoß

‘The mother ventilates again.‘

particle	non-split	Die	Mutter	mitarbeitet	wieder	
particle	split	Die	Mutter	arbeitet	wieder	mit
		‘The mother collaborates again.‘				
prefix	non-split	Die	Mutter	bearbeitet	den Antrag	
prefix	split	Die	Mutter	arbeitet	den Antrag	be
		‘The mothe edits the application.‘				
target	non-split	Der	Hobbypilot	bruchlandet	manchmal	
target	split	Der	Hobbypilot	landet	manchmal	bruch
		‘The hobby pilot sometimes does a crash landing.‘				
particle	non-split	Der	Hobbypilot	ausweicht	manchmal	
particle	split	Der	Hobbypilot	weicht	manchmal	aus
		‘The hobby pilot				
prefix	non-split	Der	Hobbypilot	entweicht	dem Hindernis	
prefix	split	Der	Hobbypilot	weicht	dem Hindernis	ent
		‘The hobby pilot avoids the obstacle.‘				
target	non-split	Die	Kundin	probesitzt	noch	
target	split	Die	Kundin	sitzt	noch	probe
		‘The client still sits on trial basis.‘				
particle	non-split	Die	Kundin	zuschaut	noch	
particle	split	Die	Kundin	schaut	noch	zu
		‘The client still watches.‘				
prefix	non-split	Die	Kundin	beschaut	die Preisliste	
prefix	split	Die	Kundin	schaut	die Preisliste	be
		‘The client looks at the price list.‘				
target	non-split	Die	Ärztin	notoperiert	oft	
target	split	Die	Ärztin	operiert	oft	not
		‘The doctor often performs emergency surgery.‘				
particle	non-split	Die	Ärztin	mitschreibt	oft	
particle	split	Die	Ärztin	schreibt	oft	mit
		‘The doctor often takes notes.‘				

prefix	non-split	Die	Ärztin	verschreibt	das Medikament	
prefix	split	Die	Ärztin	schreibt	das Medikament	ver

‘The doctor prescribes the medication.‘

target	non-split	Die	Sportlerin	eiskunstläuft	noch	
target	split	Die	Sportlerin	läuft	noch	eiskunst

‘The athelete still does figure skating.‘

particle	non-split	Die	Sportlerin	mitspielt	noch	
particle	split	Die	Sportlerin	spielt	noch	mit

‘The athelete still plays in the game.‘

prefix	non-split	Die	Sportlerin	verspielt	den Ball	
prefix	split	Die	Sportlerin	spielt	den Ball	ver

‘The athelete games away the ball.’

target	non-split	Die	Pilotin	notlandet	wieder	
target	split	Die	Pilotin	landet	wieder	not

‘The pilot does an emergency landing again.‘

particle	non-split	Die	Pilotin	abhebt	wieder	
particle	split	Die	Pilotin	hebt	wieder	ab

‘The pilot takes off again.‘

prefix	non-split	Die	Pilotin	behebt	das Problem	
prefix	split	Die	Pilotin	hebt	das Problem	be

‘The pilot solves the problem.‘

target	non-split	Die	Erstklässlerin	seilspringt	manchmal	
target	split	Die	Erstklässlerin	springt	manchmal	seil

‘The first-grader sometimes does rope skipping.‘

particle	non-split	Die	Erstklässlerin	abschreibt	manchmal	
particle	split	Die	Erstklässlerin	schreibt	manchmal	ab
prefix	non-split	Die	Erstklässlerin	beschreibt	die Geschichte	
prefix	split	Die	Erstklässlerin	schreibt	die Geschichte	be

‘The first-grader describes the story.‘

target	non-split	Die	Mechanikerin	probefährt	noch	
target	split	Die	Mechanikerin	fährt	noch	probe

‘The mechanic drives on trial basis.‘

particle	non-split	Die	Mechanikerin	aufbaut	noch	
particle	split	Die	Mechanikerin	baut	noch	auf
prefix	non-split	Die	Mechanikerin	verbaut	das Material	
prefix	split	Die	Mechanikerin	baut	das Material	ver

‘The mechanic uses the material.‘

target	non-split	Der	Direktor	fotokopiert	wieder	
target	split	Der	Direktor	kopiert	wieder	foto

‘The director photo copies again.‘

particle	non-split	Der	Direktor	zurücktritt	wieder	
particle	split	Der	Direktor	tritt	wieder	zurück
prefix	non-split	Der	Direktor	betritt	den Raum	
prefix	split	Der	Direktor	tritt	den Raum	be

‘The director enters the room.‘

target	non-split	Der	Lehrer	korrekturliest	oft	
target	split	Der	Lehrer	liest	oft	korrektur

‘The teacher often proof reads.‘

particle	non-split	Der	Lehrer	zustimmt	oft	
particle	split	Der	Lehrer	stimmt	oft	zu

‘The teacher often agrees.‘

prefix	non-split	Der	Lehrer	bestimmt	die Note	
prefix	split	Der	Lehrer	stimmt	die Note	be

‘The teacher decides the grade.‘

target	non-split	Der	Schüler	kopfrechnet	manchmal	
target	split	Der	Schüler	rechnet	manchmal	kopf

‘The pupil sometimes does mental arithmetics.‘

particle	non-split	Der	Schüler	nachsitzt	manchmal	
particle	split	Der	Schüler	sitzt	manchmal	nach

‘The pupil sometimes has detention.‘

prefix	non-split	Der	Schüler	besitzt	den Laptop

prefix	split	Der	Schüler	sitzt	den Laptop	be
		‘The pupil ownst he laptop.‘				
target	non-split	Die	Köchin	dampfgart	oft	
target	split	Die	Köchin	gart	oft	dampf
		‘The chef often steam cooks.‘				
particle	non-split	Die	Köchin	nachfragt	oft	
particle	split	Die	Köchin	fragt	oft	nach
		‘The chef often asks questions.‘				
prefix	non-split	Die	Köchin	erfragt	das Rezept	
prefix	split	Die	Köchin	fragt	das Rezept	er
		‘The chef asks fort he recipe.‘				
target	non-split	Der	Komiker	bauchredet	oft	
target	split	Der	Komiker	redet	oft	bauch
		‘The comedian often belly speaks.‘				
particle	non-split	Der	Komiker	auftritt	oft	
particle	split	Der	Komiker	tritt	oft	auf
		‘The comedian often performs.‘				
prefix	non-split	Der	Komiker	zertritt	den Krümel	
prefix	split	Der	Komiker	tritt	den Krümel	zer
		‘The comedian steps on the crumb.‘				
target	non-split	Der	Betrüger	raubkopiert	manchmal	
target	split	Der	Betrüger	kopiert	manchmal	raub
		‘The scammer sometimes pirates.‘				
particle	non-split	Der	Betrüger	einbricht	manchmal	
particle	split	Der	Betrüger	bricht	manchmal	ein
		‘The scammer sometimes breaks in.‘				
prefix	non-split	Der	Betrüger	zerbricht	die Vase	
prefix	split	Der	Betrüger	bricht	die Vase	zer
		‘The scammer breaks the vase.‘				

Appendix C

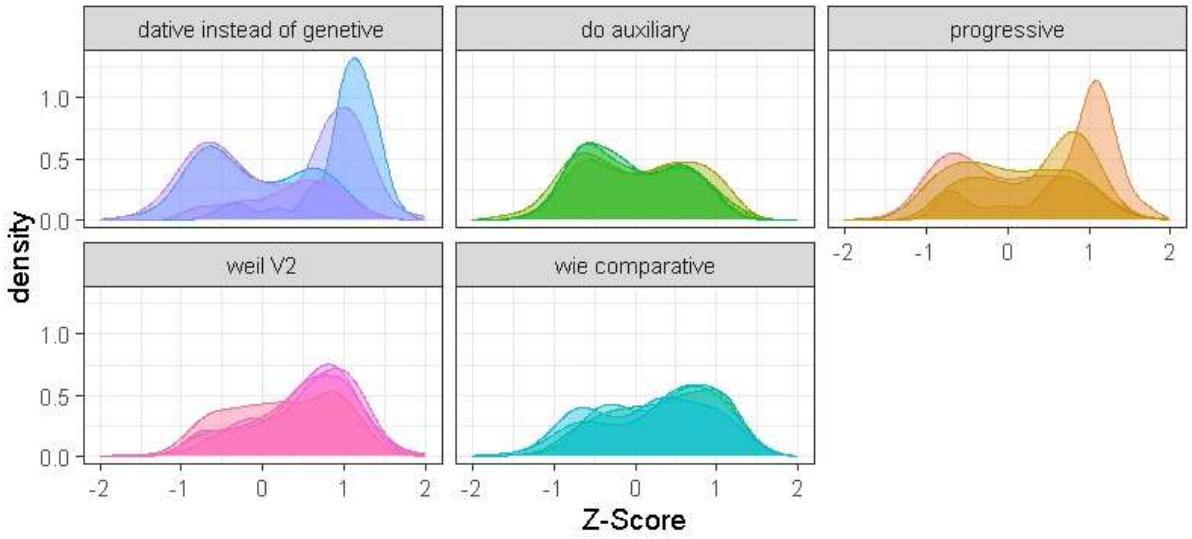
Table C1

Mean standardised and unstandardised ratings per type of norm violation filler. Standard deviatios are given in parentheses

Norm violation type	Unstandardised	Standardised
Dative instead of genetive	63.53 (34.72)	0.35 (0.81)
Do auxiliary	48.77 (30.96)	-0.03 (0.64)
Progressive	60.75 (33.21)	0.28 (0.75)
Weil V2	66.96 (28.97)	0.44 (0.63)
Wie comparative	61.73 (31.22)	0.30 (0.69)

Figure C1

Density plots of the z-score ratings for the five different norm violations



Appendix D

Model 1 – Model Selection

Table D1

List of Tested Models and their AIC scores. Selected model in bold.

Model	*df*	AIC
zscore ~ verb * context + (1 | word), REML = T	8	15487.63
zscore ~ verb * context + (1 | word) + (1 | ResponseId), REML = T	9	15453.41
zscore ~ verb * context + (1 + context | word) + (1 | ResponseId), REML = T	11	12950.02
zscore ~ verb * context + (1 + context | word) + (1 + context | ResponseId), REML = T	**13**	**12877.35**

Model 2 – Model Selection

Table B2

List of Tested Models and their AIC score. Selected model in bold.

Model	***df***	**AIC**
zscore ~ context + as.numeric(pre_s) + m_freq + Age + (1 | word), REML = F)	7	9483.51
zscore ~ context * as.numeric(pre_s) + m_freq + Age + (1 | word), REML = F)	8	9445.44
zscore ~ context * as.numeric(pre_s) * m_freq + Age + (1 | word), REML = F)	11	9394.41
Random Effects		
Model	**df**	**AIC**
zscore ~ context * as.numeric(pre_s) * m_freq + Age + (1 | word), REML = T)	11	9427.43
zscore ~ context * as.numeric(pre_s) * m_freq + Age + (1 | word) + (1 | ResponseId), REML = T)	12	9333.55
zscore ~ context * as.numeric(pre_s) * m_freq + Age + (1 + context | word) + (1 | ResponseId), REML = T)	14	7714.16
zscore ~ context * as.numeric(pre_s) * m_freq + Age + (1 + context | word) + (1 + context | ResponseId), REML = T)	**16**	**7592.41**

Appendix E

Model Criticism – Model 1

Assumption: Homoscedasticity

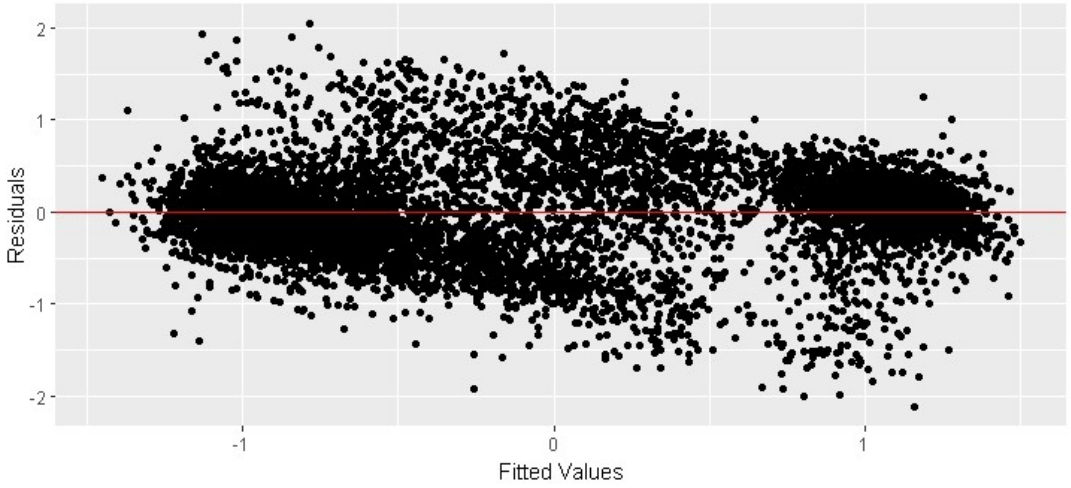
Table E1

Results of linear model predicting residuals from fitted values for Model 1

Predictor	***β***	**[95% CI]**	***SE***	***t***	***p***
Fitted values	0.004	[-0.005, 0.013]	0.004	0.87	.383

Figure E1

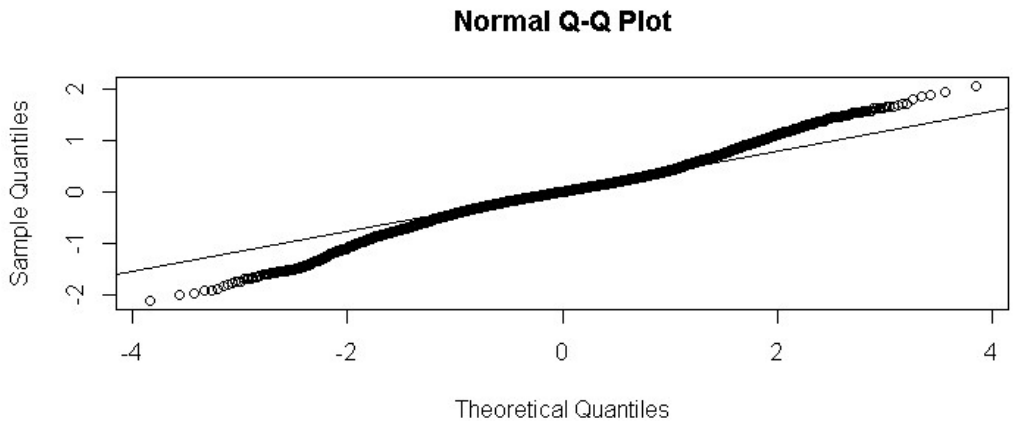
Scatter plot of residuals and fitted values of Model 1



Assumption: Normality of residuals

Figure E2

Q-Q-plot of Model 1



Assumption: Error of residuals equal to 0.

E[εi]= 2.07298e-16

Model Criticism – Model 2

Assumption: Homoscedasticity

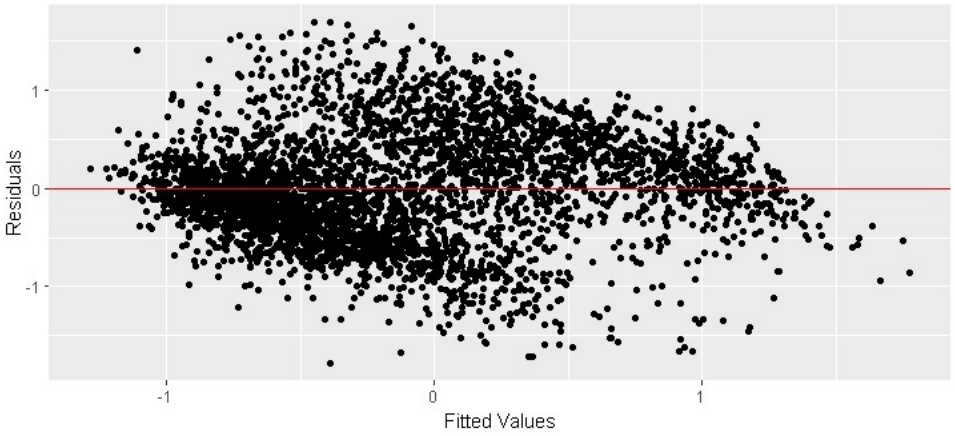
Table E2

Results of linear model predicting residuals from fitted values for Model 2

Predictor	***β***	**[95% CI]**	***SE***	***t***	***p***
Fitted values	0.079	[0.011, 0.026]	0.009	8.83	< .001

Figure E3

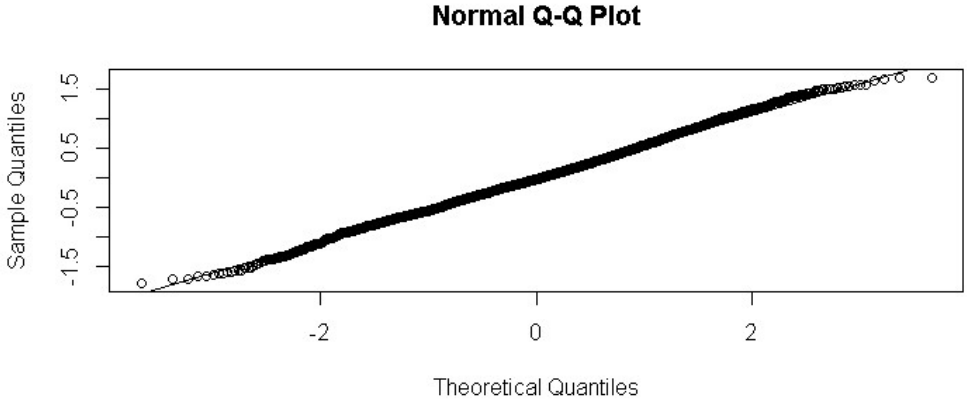
Scatter plot of residuals and fitted values of Model 2



Assumption: Normality of residuals

Figure E4

Q-Q-plot of Model 2



Assumption: Error of residuals equal to 0.

E[εi]= 3.81066e-16

Appendix F

Table F1

Silhouette scores per number of clusters of items. Selected number of cluster in bold

Number of Clusters (k)	Silhouette Score
2	0.43
3	0.44
4	0.38
5	0.40
6	**0.44**
7	0.43
8	0.40
9	0.41
10	0.39
11	0.41
12	0.41
13	0.38
14	0.35
15	0.32
16	0.29
17	0.26
18	0.24
19	0.25
20	0.23
21	0.22
22	0.22
23	0.21
24	0.19
25	0.16
26	0.15
27	0.11
28	0.08
29	0.05

Table F2

Silhouette scores per number of clusters of participants for k < 10. Selected number of clusters in bold

Number of Clusters (k)	Silhouette Score
2	0.35
3	0.34
4	0.35
5	0.36
6	0.37
7	0.35
8	0.38
9	**0.39**