

# Constructional contamination in morphology and syntax

## Four case studies

Dirk Pijpops<sup>1,2</sup> Isabeau De Smet<sup>1,2</sup> and Freek Van de Velde<sup>1</sup>

<sup>1</sup> KU Leuven | <sup>2</sup> FWO

In every-day language use, two or more structurally unrelated constructions may occasionally give rise to strings that look very similar on the surface. As a result of this superficial resemblance, a subset of instances of one of these constructions may deviate in the probabilistic preference for either of several possible formal variants. This effect is called ‘constructional contamination’, and was introduced in Pijpops & Van de Velde (2016). Constructional contamination bears testimony to the hypothesis that language users do not always execute a full parse of the utterances they interpret, but instead often rely on ‘shallow parsing’ and the storage of large, unanalyzed chunks of language in memory, as proposed in Ferreira, Bailey, & Ferraro (2002), Ferreira & Patson (2007), and Dąbrowska (2014).

Pijpops & Van de Velde (2016) investigated a single case study in depth, namely the Dutch partitive genitive. This case study is reviewed, and three new case studies are added, namely the competition between long and bare infinitives, word order variation in verbal clusters, and preterite formation. We find evidence of constructional contamination in all case studies, albeit in varying degrees. This indicates that constructional contamination is not a particularity of the Dutch partitive genitive but appears to be more widespread, affecting both morphology and syntax. Furthermore, we distinguish between two forms of constructional contamination, viz. first degree and second degree contamination, with first degree contamination producing greater effects than second degree contamination.

**Keywords:** constructional contamination, corpus, Dutch, alternation, partitive genitive, long infinitive, bare infinitive, verbal clusters, weak verbs, strong verbs

## 1. Introduction

Much of linguistics has been strictly concerned with describing the language ‘system’ and how this system determines our day-to-day language usage (Geeraerts 2010:73–77). Under this view, language users are assumed to process language only by performing operations upon the grammatical structures that are part of the system and underlie their utterances (Gibson 1991; Phillips 1996; Weinberg 1993). In recent decades, however, usage-based linguistics has largely reversed this perspective, focusing not on how the system of grammatical structures determines our language use but, rather, how language usage is constantly changing and shaping the system (Geeraerts 2010:83; for examples see Hopper 1987; Bybee 2010; Diessel 2015). The development and design of the system is here considered to be the cumulative result of millions of usage events (Beckner et al. 2009; Larsen-Freeman & Cameron 2008; Steels 2000; van Trijp 2008). Within usage-based linguistics, it has also been argued that language users do not always construct detailed syntactic parses while processing but may also store larger, unanalyzed language chunks in memory, which can in turn be accessed as ready-mades during production (Dąbrowska 2014; Ferreira et al. 2002). From this view it follows that chunks of language that are superficially identical, yet structurally different, should occasionally interfere with one another during processing, since these structural differences are not consistently uncovered by language users.

One way of how such interference may occur has been described by Pijpops & Van de Velde (2016), who have called it ‘constructional contamination’. A formal definition of this mechanism can be found in section 2 but here we give an introductory example. The Dutch partitive genitive construction comes in two formal variants: one with and one without *-s* ending, as can be seen in (1–2) (Pijpops & Van de Velde 2018).<sup>1</sup> When particular partitive genitives, such as the ones with the adjective *verkeerd* ‘wrong’, appear in the variant without *-s*, they show a superficial resemblance to another construction, namely, combinations with adverbs. In (2b) and (3), one may see the same superficial string *iets verkeerd* ‘something wrong(ly)’, although the underlying syntactic structure of both utterances is markedly different and completely unrelated. That is, *verkeerd* in (2b) is an adjective modifying the pronoun *iets* ‘something’, and together, they compose a noun phrase in a partitive genitive construction, while in (3), *verkeerd* is an adverb modifying the verb *geïnterpreteerd* ‘interpreted’. Such superficial resemblance is not present for the variant with *-s* ending in (2a) because Dutch adverbs cannot gram-

---

1. The *-s* ending is glossed as *-GEN*, because it historically derives from genitive inflection. The genitive case has long disappeared from the Dutch language, however, and the ending is now best characterized as an isolated suffix (Haeseryn et al. 1997:863).

matically receive an *-s* suffix in such instances. The superficial resemblance is also not present for other adjectives like *zinnig* 'sensible' because *zinnig* is only rarely used as an adverb in combinations such as (3) (Pijpops & Van de Velde 2016: 570).

(1) PARTITIVE GENITIVE:

a. *Ik heb iets zinnig-s gelezen.*

b. *Ik heb iets zinnig gelezen.*

I have something sensible(-GEN) read

'I have read something sensible.'

(2) PARTITIVE GENITIVE:

a. *Ik heb iets verkeerd-s gegeten.*

b. *Ik heb iets verkeerd gegeten.*

I have something wrong(-GEN) gegeten

'I have eaten something wrong.'

(3) COMBINATION WITH ADVERB:

*Ik heb iets verkeerd geïnterpreteerd.*

I have something wrongly interpreted

'I have interpreted something in the wrong way.'

The frequent occurrence of the string *iets verkeerd* in combinations such as (3) reinforces the representation of this chunk in the memory of language users.<sup>2</sup> When these language users then need to form a partitive genitive with the adjective *verkeerd* 'wrong', such as in (2), they will have developed a preference for the form *iets verkeerd*, i.e. the variant without *-s*. Meanwhile, this will not have happened for the partitive genitives with *zinnig* 'sensible'. As a result, the variant without *-s* will be more prevalent among partitive genitives with *verkeerd* such as (2) than among partitive genitives with *zinnig* such as (1). This is constructional contamination.

The mechanism of constructional contamination then crucially depends upon a combination of two asymmetries. The first is a Type 2 paradigmatic asymmetry where one meaning corresponds to many forms, such as both formal variants of the partitive genitive construction in (1) and (2) (cf. Koutsoukos et al. in this issue on different types of asymmetries and mismatches). The second is a Type 1 paradigmatic asymmetry where the same phonological form can be interpreted in multiple ways, such as *iets verkeerd* in (2b) and (3).

Pijpops & Van de Velde (2016) present an in-depth investigation into the Dutch partitive genitive as a single case study of constructional contamination. In

2. There is strong evidence that language users indeed work on such chunks even if they are inconsistent with the global syntactic structure, cf. research on *local syntactic coherences* in, among others, Tabor et al. (2004), Konieczny (2005), and Konieczny et al. (2009).

the present article, we cast our nets wider and take under scrutiny three further case studies from syntax and morphology. Our goal here is to show that constructional contamination is not an idiosyncratic curiosity of the Dutch partitive genitive alternation but, rather, constitutes a fairly wide-spread phenomenon.

The next section is theoretical in nature and defines constructional contamination in an abstract way. There, we also further develop the notion of constructional contamination by distinguishing between two forms of it, namely, first degree and second degree contamination. Section 3 forms the bulk of the article, with each of its subsections presenting a separate case study. First, we review the results of Pijpops & Van de Velde (2016) on the partitive genitive. Second, we discuss the competition between the bare infinitive and the long infinitive as complements of auxiliaries. Third, we deal with word order variation in verbal clusters. Fourth, we turn to the competition between the weak and strong preterite inflection, building further on research by Vosters (2012). All case studies are taken from Dutch. The ease with which we could already identify several potential case studies in a single language can be taken as an optimistic sign that many more are to be discovered in other languages. Still, we leave such enterprises to experts of these other languages, at least for now. Finally, section 4 summarizes the results, discusses avenues for further research on constructional contamination and explains why constructional contamination is relevant for research on language processing, alternation studies, and Construction Grammar theory.

## 2. The mechanism and effect of constructional contamination

Constructional contamination is best understood as a specific type of analogical interference, on a par with other types such as four-part analogy, paradigmatic levelling etc. The term is used both for the mechanism causing the interference and the effect resulting from it. We begin by characterizing the mechanism and then describe how one may observe the effect in language corpora.

In what follows, we will use the following abbreviations. We use upper case letters to refer to constructions, i.e. constructional schemas, and the corresponding lower case letters to refer to subsets of instances, i.e. constructs, of those constructions.

- The construction under scrutiny is named the target construction *T*, e.g. the partitive genitive.
  - The first variant of the target construction is named *TX*, e.g. the partitive genitive without -s ending.
  - The second variant of the target construction is named *TY*, e.g. the partitive genitive with -s ending.

- There is another construction, some of whose instances show superficial resemblance to the target construction. This construction is named the contaminating construction  $C$ , e.g. the combinations with adverbs.
  - The subset of instances of  $C$  that show superficial resemblance to  $T$  is named  $c_i$ , e.g. *ik heb iets verkeerd geïnterpreteerd* ‘I have interpreted something in the wrong way’.
- One subset of instances of  $T$  is named  $t_o$ . These instances never show superficial resemblance to  $C$ , neither in the variant  $TX$ , nor in the variant  $TY$ , e.g. *iets leuk(s)* ‘something fun’, *iets zinnig(s)* ‘something sensible’, etc.
- Another subset of instances of  $T$  is named  $t_i$ , e.g. *iets verkeerd(s)* ‘something wrong’, *iets goed(s)* ‘something good’, etc.
  - When these instances  $t_i$  appear in the form of  $TX$ , they are named  $tx_i$ . The forms  $tx_i$  show a superficial resemblance to the contaminating construction  $C$ , e.g. *iets verkeerd* ‘something wrong’, *iets goed* ‘something good’, etc.
  - When these instances  $t_i$  appear in the form of  $TY$ , they are named  $ty_i$ . The forms  $ty_i$  do not show superficial resemblance to the contaminating construction  $C$ , e.g. *iets verkeerds* ‘something wrong’, *iets goeds* ‘something good’, etc.

We can now define the mechanism of constructional contamination. The mechanism crucially depends on the chunking method of language processing, as described in Ferreira & Patson (2007) and Dąbrowska (2014). For it to kick into action, the following two conditions need to be met.

Condition A. One-meaning-multiple-forms asymmetry: the target construction  $T$  displays some form of formal alternation between its variants  $TX$  and  $TY$ .<sup>3</sup>

Condition B. One-form-multiple-meaning asymmetry: the contaminating construction  $C$  has a subset of instances  $c_i$  that are superficially (near)

---

3. Alternatively, a construction is in competition with another construction that is distinct in form. In what follows, we describe constructional contamination as an effect that may arise in the competition of two formal variants of a single construction but it may also arise in a competition of two constructions that are formally distinct, yet semantically similar enough to be interchangeable, such as the ditransitive and prepositional dative construction (Bresnan et al. 2007; Coleman 2009; Röthlisberger et al. 2017). In that case, the same mechanism would be at work, *mutatis mutandis*. For example, the ditransitive construction would correspond to  $TX$  and the prepositional dative construction to  $TY$ . In either case, the variation may well be multifactorially determined. This means that the influence of constructional contamination may simply be one of various alternation factors. For the Dutch partitive genitive, it was even shown to be the prime factor determining the alternation (Pijpops & Van de Velde 2016: 567).

identical to a subset of instances  $t_i$  of construction  $T$ , when these instances take the form  $tx_i$  of variant  $TX$ .

Now, the language user repeatedly hears the instances  $c_i$ . If he or she employs exemplar chunking, s/he will, at least occasionally, store an instance of  $c_i$  as an unanalyzed chunk of language. For example, upon hearing the utterance *ik heb iets verkeerd geïnterpreteerd* ‘I have interpreted something in the wrong way’, the chunk *iets verkeerd* is being stored. The same will also happen for the forms  $tx_i$ . For example, upon hearing the utterance *ik heb iets verkeerd gegeten* ‘I have eaten something wrong’, the chunk *iets verkeerd* is being stored. Because both memory representations constitute unanalyzed chunks of language, they reinforce one another. In other words,  $tx_i$  benefits from its superficial resemblance to  $c_i$  in that hearing  $c_i$  increases the cognitive entrenchment of the representation of  $tx_i$ . In the case of the partitive genitive, this would mean that hearing *iets verkeerd* in *ik heb iets verkeerd geïnterpreteerd* ‘I have interpreted something in the wrong way’ reinforces the cognitive entrenchment of *iets verkeerd* ‘something wrong’, i.e. the partitive genitive form without *-s* ending.<sup>4</sup>

Meanwhile, the same does not occur for the competing forms  $ty_i$ , because these forms do not show superficial resemblance to any instances of the contaminating construction  $C$ . For example, the same does not happen for the form *iets verkeerd*s because the chunk *iets verkeerd*s never appears as an adverbial combination.<sup>5</sup> One possible reason for this is that  $C$  simply cannot grammatically appear in an instance that resembles  $ty_i$ . This is the case for *iets verkeerd*s: Dutch adverbs cannot grammatically receive an *-s* ending in such combinations. Another possible reason is that, while  $C$  can in principle appear in an instance that resembles  $ty_i$ , it does so in practice far more rarely than in an instance that resembles  $tx_i$ . To sum up, the forms  $tx_i$  receive aid from the similar forms  $c_i$ , while such aid is not available, or less so, to their competitors  $ty_i$ .

We therefore expect the instances  $t_i$  to exhibit a skew in their formal realization, with a probabilistic preference for variant  $TX$ . This probabilistic skew would be in comparison to the instances  $t_o$ , where neither variant exhibits any superficial resemblance to the contaminating construction  $C$ . This is the effect of constructional contamination. In particular, it is the effect of first-degree constructional contamination.

We now turn to second degree constructional contamination. For this, we introduce one last abbreviation.

4. We talk about *hearing* here but the same would hold for other forms of language comprehension, such as reading written language or viewing sign language.

5. Disregarding here the effect of constructional contamination in the opposite direction (see Pijpops & Van de Velde 2016: 545, Fn. 4 for this effect in partitive genitives, and section 3.3 in the present article for this effect in verbal clusters).

- Another subset of instances of  $T$  are named  $t_2$ . These instances do not show direct superficial overlap with the instances  $c_i$ , but they do show superficial resemblance to the affected instances  $t_i$ . For the partitive genitive, we have yet to identify such instances, so we cannot yet give any examples. Examples will be given in section 3.2, however.

Once the instances  $t_i$  are affected, the effect can then spread to instances  $t_2$ . This can happen through the same mechanism as just described, with  $t_i$  as the contaminating instances, and  $t_2$  as the target instances. We call this second degree constructional contamination. Naturally, we expect the effect of first degree constructional contamination to be larger than that of second degree contamination. The reason is that the second degree contamination is dependent on the first degree contamination in the same way as an offshoot branch is dependent upon a parent tree.

The instances  $t_i$  that show direct resemblance to the contaminating construction should thus exhibit the strongest preference for the variant  $TX$ , followed by the instances  $t_2$  that resemble  $t_i$ , but show no direct overlap with instances of the contaminating construction. Finally, the instances  $t_o$  that show no resemblance whatsoever to the contaminating construction or to the affected instances should also exhibit the least propensity for the variant  $TX$ . This constitutes a prediction that is easily testable in the corpora of natural language use. In the next section, we do just that. The first case study only deals with first degree contamination; second degree contamination is added in the second case study.

### 3. Four case studies of constructional contamination

#### 3.1 The partitive genitive

The partitive genitive construction is a combination of a quantifier, in the form of an indefinite pronoun or numeral, and an adjectival phrase, which together build a nominal phrase. Corpus examples can be found in (4–7).<sup>6</sup> The construction runs largely parallel to its counterpart in English, e.g. *something interesting*, the most notable exception being that the English construction never employs an *-s* ending. The details of its structural analysis in Dutch, such as which of its constituents is its syntactic head, or what is the nature of the relation between quantifier and adjectival phrase (e.g. modifying or predicative), are still being debated but these are not at issue here (see Broekhuis & Strang 1996; Kester

---

6. All examples given in this section are taken from the ConDiv corpus (Grondelaers et al. 2000).

1996; Hoeksema 1998; Booij 2010:223–228; Pijpops & Van de Velde 2015; or for an overview, Pijpops & Van de Velde 2018:103–106).

- (4) *Iemand nog iets interessants te melden?*  
 someone still something interesting to report  
 ‘Does anyone still have something interesting to report?’
- (5) *Uitslapen en beetje met ouwe atari klooien... niet veel speciaals*  
 sleep.out and bit with old Atari bungle not much special  
 ‘Sleep in and play around with an old Atari, nothing special really.’
- (6) *ok hier nog mensen die iets interessant te melden hebben*  
 ok here still people that something interesting to report have  
 ‘Ok, are there still any people here that have something interesting to report?’
- (7) *Ik ben ook eens gaan kijken, veel speciaal is er inderdaad niet te vinden*  
 I am also once go look much special is there indeed not to find  
 ‘I have also gone and taken a look, there is indeed not much of interest to be found there.’

Of more relevance to the current issue than the internal syntax of the construction is the fact that the partitive genitive meets both conditions of constructional contamination. As for condition A, the partitive genitive displays formal variation between a variant with an *-s* ending on the adjective, as in (4–5) and a variant without an *-s* ending, as in (6–7). As for condition B, there are at least two other, unrelated constructions whose instances are often superficially similar to instances of the variant without the *-s* ending. We have already introduced the combinations with adverbs, as in (8) – compare to (9); another source of contamination are subject complements shown in (10) or other predicative constructions, in comparison to (11). The partitive genitive and both contaminating constructions are simple run-of-the-mill language structures of Dutch.

- (8) COMBINATION WITH ADVERB:  
*... dat iets verkeerd geïnterpreteerd wordt?*  
 that something wrongly interpreted gets  
 ‘...that something gets wrongly interpreted?’
- (9) PARTITIVE GENITIVE:  
*In begin van de week iets verkeerd gegeten*  
 in beginning of the week something wrong eaten  
 ‘I’ve eaten something wrong at the beginning of this week.’



## (10) SUBJECT COMPLEMENT:

*Clakkeloos aannemen dat iets goed of fout is*  
 unquestionably accepting that something good or wrong is  
 'Unquestionably accepting that something is good or wrong.'

## (11) PARTITIVE GENITIVE:

*Dan ben je verzekerd van iets goed*  
 then are you assured of something good  
 'Then you're sure to get something good.'

The superficial resemblance is not present for all instances of the partitive genitive, or at least not to the same extent. The reason is that instances such as (8) and (10) are particularly frequent with the so-called assessment adverbs/adjectives, such as *verkeerd* 'wrong', *goed* 'good', *beter* 'better', and *fout* 'incorrect'. They are markedly less frequent with other adjectives, such as *zinnig* 'sensible' or *leuk* 'fun'. In Pijpops & Van de Velde (2016), we tested four metrics that quantified for each adjective, or for each combination of a quantifier and an adjective, how often the same string without the -s ending appeared in a construction that was not a partitive genitive. These metrics were then used to predict the propensity for -s omission among strictly unambiguous partitive genitives, when controlling for all other known factors to determine the alternation. All of them showed positive correlations. This confirms the prediction of first degree constructional contamination, as formulated in section 2.

### 3.2 Long and bare infinitives

Like in English, auxiliaries in Dutch can be classified according to the type of complement they take.<sup>7</sup> Some take a participle (more on this in section 3.3), while others take an infinitival complement. Within the latter group, there are auxiliaries that occur with a bare infinitive and other auxiliaries that occur with a long infinitive (or: *to*-infinitive). Examples are given in (12–13). The structures in Dutch and English are exactly parallel.

## (12) Bare infinitive:

*Dat moet Ø/\*te werken.*  
 'That must work.'

---

7. Some scholars use a more restricted notion of 'auxiliary', reserving the term, for instance, only for the perfective auxiliaries. We take a broad view here, including verbs like *beginnen* 'begin' or *leren* 'teach/learn'. The reason is that these verbs can form 'clusters', which can be seen as a reliable formal criterion for an auxiliary status in Dutch, and that auxiliaries form a continuum (Bolinger 1980; Heine 1993; van der Horst 2008: 873–896; Coupé 2015: 13–18).

- (13) Long infinitive:  
*Dat lijkt te/\*Ø werken.*  
 ‘That seems to work.’

For many auxiliaries, the choice between a bare and long infinitive is fixed: core modals like *moeten* ‘must’ invariably go with the bare infinitive whereas other auxiliaries, like *lijken* ‘seem’, go with the long infinitive.<sup>8</sup> Concerning the three so-called posture auxiliaries *liggen* ‘lie’, *staan* ‘stand’, and *zitten* ‘sit’, which express durative aspect, reference grammars of Dutch claim that only the long infinitive is grammatically possible if the auxiliary is finite, as in (14) (Haeseryn et al. 1997: 970–974; Van Bart et al. 1998: 54; Klooster 2001: 61–63).<sup>9</sup> Meanwhile, if the auxiliary is itself an infinitive, both bare and long infinitive are grammatical. Such a context occurs in the so-called Infinitivus Pro Participio (IPP, also called *Ersatz-infinitiv*). The IPP-construction is unknown in English but occurs in Dutch and German: when used in the perfect, auxiliaries may occur in the infinitive instead of in the past participle (see Den Besten & Edmonson 1983; Schmid 2005; Coupé 2015: 48–51; Augustinus & Van Eynde 2017). An example of IPP with the bare infinitive is given in (15).

- (14) [FINITE POSTURE AUX + INFINITIVE]: bare infinitive is ungrammatical  
*Hij zit te slapen.*  
 he sits to sleep  
 ‘He is sleeping.’
- (15) IPP-construction: bare infinitive is grammatical and frequent  
*Hij heeft de hele les zitten slapen.*  
 he has the entire class sit sleep  
 ‘He has been sleeping throughout the entire class.’

However, there is one peculiar exception to the rule that finite posture auxiliaries cannot combine with the bare infinitive. This exception states that if the auxiliary is in the indicative of the simple present plural in a subordinate clause, the bare infinitive is possible as well, as in (16) (Haeseryn et al. 1997: 970; Klooster 2001: 61).

- (16) [FINITE AUX + INFINITIVE]: bare infinitive exceptionally grammatical  
*Als die jongens de hele les zitten slapen, zullen ze niet veel opsteken*  
 if those boys the entire class sit sleep will they not much pick.up  
 ‘If those boys are sleeping throughout the entire class, then they won’t learn much’  
 (example taken from Haeseryn et al. 1997: 970)

8. Diachronically, some shifts in the distribution of bare vs. long infinitives can be discerned (see van der Horst 2008; Van de Velde 2017).

9. For more information on these three auxiliaries, see Lemmens (2005).

This exception is hard to explain by looking at the underlying structure of the sentence but becomes rather obvious when looking at the surface string. If the auxiliary is indicative of simple present plural and part of a subordinate clause, as in (16), then the string *zitten slapen* is superficially identical to the string *zitten slapen* in the IPP-construction – compare with (15). In other words, because of the frequent occurrence of the string *zitten slapen*, without the intervening *te*, in the IPP-construction in (15), this string becomes entrenched in memory. When language users then need to produce a sentence as in (16), the normally ungrammatical combination of a finite posture auxiliary and a bare infinitive suddenly becomes acceptable. This presents a case of constructional contamination, with the IPP-construction as the contaminating construction and the [FINITE POSTURE AUX + INFINITIVE] construction as the target construction.

In particular, instances such as (16) are affected by first degree constructional contamination, since they exhibit superficial formal identity to the contaminating construction. In turn, the contaminating influence may spread further from these instances, affecting those that resemble (16), such as (17). This would qualify as second degree contamination, since instances such as (17) do not exhibit direct superficial identity with the contaminating IPP-construction. We therefore expect that these instances would also exhibit the – allegedly ungrammatical – bare infinitive, but to a far lesser degree than instances like (16).

- (17) [FINITE POSTURE AUX + INFINITIVE]: bare infinitive normally ungrammatical  
*Als die jongens de hele les zaten slapen, hebben ze niet veel*  
 if those boys the entire class sat sleep have they not much  
*opgestoken*  
 pick.up  
 ‘If those boys have been sleeping throughout the entire class, they haven’t  
 learned much’

This allows us to test the prediction spelled out in section 2. For the present case study, the abbreviations introduced in section 2 correspond to the following constructions and constructs.

- The target construction *T*: [FINITE POSTURE AUX + INFINITIVE]
  - The first variant of the target construction *TX*: the bare infinitive
  - The second variant of the target construction *TY*: the long infinitive
- The contaminating construction *C*: IPP-construction
- The subset of instances *t<sub>o</sub>* that never show superficial resemblance to the IPP-construction. These are the instances where the finite auxiliary and the infinitive are not in a verb-final clause, such as (18).

- The subset of instances  $t_1$  that exhibit superficial resemblance to the IPP-construction, when they appear in the form of the bare infinitive. These are the instances where a finite verb and an infinitive are part of a verb-final clause, and the finite verb is indicative simple present plural and thus formally identical to its infinitival form, such as (16), in comparison to (15).
- The subset of instances  $t_2$  that do not show direct superficial resemblance with the IPP-construction, but do exhibit near-identity to the affected instances  $t_1$ . These are the instances where a finite auxiliary and an infinitive are part of a verb-final clause but the finite verb is not homophonous with its infinitival form, such as (17).

- (18) *De jongen zit verdorie al heel de les \*(te) slapen.*  
 the boy sits darned already whole the class to sleep  
 ‘Darned, the boy has already been sleeping throughout the entire class.’

Group  $t_1$  is theorized to be affected by first-degree constructional contamination, group  $t_2$  by second-degree constructional contamination, and group  $t_0$  is hypothesized not to be affected at all. As such, we expect group  $t_1$  to exhibit the comparatively strongest predilection for the bare infinitive, followed by group  $t_2$  and finally by group  $t_0$ .

To test this prediction, we made use of the SoNaR corpus of written Dutch (Oostdijk et al. 2013). From this corpus, all instances were extracted in which a finite posture auxiliary was part of a verb-final clause and was contiguous with either a bare or a long infinitive, as in (16) and (17), as well as all instances where the finite verb was part of a verb-second or verb-initial clause and preceded a bare or long infinitive with maximally five intervening words, as in (18).<sup>10</sup>

This yielded 28,788 instances in total, of which 26,022 exhibited a long infinitive and 2,766 exhibited a bare infinitive. 2,766 instances may seem like an awful lot for a variant that is either a peculiar exception or plainly ungrammatical and in fact, it is. The large majority of these 2,766 bare infinitive instances constituted false positives, such as (19). We manually went through these instances, with each of the three authors checking 922 of them (that is, one third). In this way, we were able to identify 7 instances of the bare infinitive in group  $t_1$ , 3 instances of the bare infinitive in group  $t_2$ , and 1 instance of the bare infinitive in group  $t_0$ . All of these instances are listed below.

10. We could not demand from the query that both the finite verb and the infinitive were part of the same clause because sentences like (14) and (15) with the bare infinitive would be incorrectly parsed by the corpus' automatic syntactic parser, Alpino (van Noord 2006). The reason is that the parser was not built to handle such instances since they are considered ungrammatical. Clauses that could not be classified by our algorithm as verb-initial, verb-second, or verb-final were not taken up.

- (19) *'t Pleksken staat op instorten.*  
*'t Pleksken stands on collapse*  
*'t Pleksken is on the verge of collapsing.'*

It is not the case that the members of group  $t_1$  are simply more frequent than those of group  $t_2$  or group  $t_o$ . The SoNaR corpus yielded only 2,622 instances of the long infinitive of group  $t_1$ , while it yielded 11,978 of group  $t_2$  and 13,576 of group  $t_o$ . If we were to find proportionally the same amount of bare infinitives in group  $t_2$  and  $t_o$  as we found in group  $t_1$ , we would expect, respectively, 31 and 36 bare infinitives. This difference between group  $t_1$  and  $t_2$ , and between  $t_1$  and  $t_o$  is significant, both with  $p < 0.001$  (Fisher's exact test). In other words, there are significantly more cases of the bare infinitive in group  $t_1$ . The raw numbers are rather low, as can be expected from ungrammatical instances, but this still confirms our hypothesis.

Note that all instances of the bare infinitive in group  $t_2$  contain a past plural of the finite verb, which has an inflectional *-en* ending that is homonymous with the infinitival ending. This means that although their superficial strings are not formally identical to the contaminating IPP-construction, there is still considerable formal resemblance.

#### Group $t_1$

- (20) ... *dat ze zitten roepen en tieren als er een doelpunt gescoord wordt.*  
 that they sit shout and yell if there a goal scored is  
 '... that they are shouting and yelling if a goal has been scored.'
- (21) *We sturen de milieu en andere problemen door naar de volgende*  
 we send the environment and other problems on to the next  
*generaties, dat deze maar in onze smeerlapperij zitten zitten*  
 generations that those just in our filth sit sit  
 'We simply pass on the environmental problems and other problems to the next generations, so they will be stuck in our filth.'
- (22) *Heerlijk aan het prutsen en klussen op de dingen die al jaren liggen*  
 delightfully on the tinker and fix on the things that already years lie  
*wachten*  
 wait  
 'Delightfully tinkering with and fixing things that have already been waiting to be done for years.'
- (23) *alle allen die sopperturke die zo straf zitten doen*  
 well only those show.off.Turks that so tough sit do  
 'Well, only those Turks who are acting all tough.'

- (24) *veel beter dan dieje stom hollanders die om de 5 seconden om geld  
much better than those stupid Dutchmen that after the 5 seconds for money  
zitten bedelen*  
sit beg  
'Much better than those stupid Dutchmen, who are begging for money every  
five seconds.'
- (25) *neeje ik meen het, ik weet niet of an dat tof vind da ze nu nog  
no I mean it I know not whether An that ok finds that they now still  
steeds naar die link zitte zien*<sup>11</sup>  
always to that link sit see  
'No, I mean it, I don't know whether An thinks it's ok that they are still looking  
at that link.'
- (26) *de 5 gratis gidsen die daar op de lezers liggen wachten*  
the 5 free guides that there on the readers lie wait  
'The 5 free guidebooks that are awaiting the readers over there.'

Group  $t_2$ 

- (27) *... dat veel mensen die genieten van een leefloon bijvoorbeeld  
that many people that enjoy of a social.security.benefit for.example  
al voor zeven uur 's ochtends aan een bushalte stonden wachten*  
already before seven hour the morning at a bus.stop stood wait  
'... that many people who live on social welfare could be seen waiting at the  
bus stop already before 7am.'
- (28) *We kauwden op de peperdure amandelen terwijl we zaten wachten tot  
we chewed on the very.expensive almonds while we sat wait until  
onze vlucht zou worden afgeroepen.*  
our flight would be called.out  
'We were chewing on the very expensive almonds while we were waiting for  
our flight to be called.'
- (29) *Twintig mensen die zaten eten ...*  
twenty people that sat eat  
'Twenty people who were eating ...'

Group  $t_0$ 

- (30) *Hij staat rustig de foto's bekijken.*  
he stands calmly the pictures look.at  
'He is calmly looking at the pictures.'

---

11. From the context, it was clear that *ze* (either 'she' or 'they') was plural ('they'). This means that *zitte* 'sit' is plural too, and not *zit te* written as one word.

### 3.3 Verbal clusters

Let us now proceed to our third case of constructional contamination, which involves word order in the verbal cluster. To understand what is going on, a few general notes on word order in the Dutch clause are needed.

Dutch word order is almost identical to German word order and can insightfully be described by the so-called topological approach with a bipolar structure (*Klammerstruktur*) (Zifonun et al. 1997:1498). The basic idea is that the clause is centered around two ‘poles’ (or *Klammern*), as visualized in Table 1. Main clauses such as (31) have the finite verb at the first pole, which is preceded by the pre-field. This pre-field contains at most a single constituent, whence the term ‘verb-second’ (V2). All other non-finite verbs go to the second pole, which is also called the verbal end group. Subordinate clauses like (32) work in the same way, except that the first pole is typically occupied by a conjunction and all verbs, including the finite verb, take up a position at the second pole.<sup>12</sup>

The second pole is where verbs can cluster. Restricting ourselves to combinations of an auxiliary and a non-finite verbal complement of the auxiliary, clustering happens under two conditions. The first condition corresponds to subordinate clauses like (32), where the auxiliary *heb* ‘have’ is finite. The second condition corresponds to main clauses like (33), where the auxiliary *hebben* ‘have’ is non-finite.

Table 1. The bipolar structure of Dutch clauses

	Prefield	1st verbal pole	Midfield	2nd verbal pole
(31) Main clause:	<i>Ik</i> I ‘I have seen your mother.’	<i>heb</i> have	<i>je moeder</i> your mother	<i>gezien.</i> seen
(32) Subordinate clause with finite auxiliary <i>heb</i> ‘have’:	∅ ‘...that I have seen your mother.’	<i>dat</i> that	<i>ik je moeder</i> I your mother	<i>heb gezien.</i> have seen
(33) Main clause with non-finite auxiliary <i>hebben</i> ‘have’:	<i>Ik</i> I ‘I must have seen your mother’	<i>moet</i> must	<i>je moeder</i> your mother	<i>hebben gezien.</i> have seen

Dutch verbal clusters have been the subject of a wide array of scholarly studies, for a number of reasons. One of them is the word order variation they display: the auxiliary and the participle can occur in two variants, namely, [AUX + PARTICIPLE] and [PARTICIPLE + AUX]. The first variant is called the ‘red’ word order and

12. Readers who are unfamiliar with the basics of Germanic word order and may be confused by this very short characterization of Dutch clause structure are advised to consult easy-access explanations provided in König & Gast (2009: Chapter 10) or Zwart (2011: 26).

the second variant is called the ‘green’ word order, hence the colors in Table 2 for (34) and (35), respectively.<sup>13</sup> The word order is often considered to be ‘semi-free’, meaning that both variants are interchangeable but the choice between them is not entirely unpredictable, as it is responsive to a considerable number of sociolectal and morphosyntactic features (see De Sutter 2005 for a book-length account, with numerous references to other studies, and Bloem et al. 2014 for a follow-up study). Moreover, the relative frequency of the word order variants has changed over time, in a back-and-forth way (Coupé 2015; Coussé 2008). Despite all this research, predicting the occurrence of red and green word order is still elusive and additional factors have been proposed in recent years (Bloem et al. 2015, 2017; Bloem 2016).

Table 2. Both word order variants of verbal clusters of the 2nd verbal pole

	Prefield	1st verbal pole	Midfield	2nd verbal pole
(34) The red order variant	Ø	<i>dat</i>	<i>ik jouw moeder</i>	<i>heb gezien.</i>
[AUX + PARTICIPLE]:		that	I your mother	have seen
		‘...that I have seen your mother.’		
(35) The green order variant	Ø	<i>dat</i>	<i>ik jouw moeder</i>	<i>gezien heb.</i>
[PARTICIPLE + AUX]		that	I your mother	seen have
		‘...that I have seen your mother.’		

As we will argue, the choice for the variants is also influenced (though of course not fully determined) by constructional contamination. To understand how this works, we look at the verbal clusters with the auxiliary *zijn* ‘be’, such as in (36) in Table 3. This auxiliary is used in the formation of the passive of the (plu-)perfect, as in (37), and in the formation of the active (plu-)perfect of unaccusative intransitives. On the surface, such clauses resemble instances with predicative adjectives in subject complement function with the copula *zijn*, as in (37–38) in Table 3. This is a result of the categorial ambiguity of participles, which can be either verbal, as in (36), or adjectival, as in (37). In fact, in the absence of contextual disambiguation cues, like a *door*-PP expressing the demoted agent of the passive, as in (35), both readings may be available (see Coussé 2011).<sup>14</sup>

13. This convention – an idiosyncrasy of the Dutch grammar tradition – goes back to the work of Pauwels (1953), who used these colours to draw dialectal maps of the geographic distribution of the two variants. Other names for the red order are the ‘1–2 order’ or the ‘ascending order’, and for the green order the ‘2–1 order’ or ‘the descending order’.

14. To be fair, this cue is not watertight, as adjectives can be accompanied by *door*-PPs as well: *De appel is bevroren/hard door de vorst* ‘The apple is frozen/hard because of the frost’.



**Table 3.** A target instance of the verbal clusters with the auxiliary *zijn* ‘be’ and two instances of the contaminating construction [ADJECTIVE + COPULA], the first of which superficially resembles the target instance

	Prefield	1st verbal pole	Midfield	2nd verbal pole
(36) Target instance of the verbal clusters:	Ø	<i>dat</i> that	<i>ik door haar</i> I by her	<i>vermoeid ben.</i> tired am
		‘...that she has been tiring me out.’		
(37) Contaminating construction: [ADJECTIVE + COPULA]	Ø	<i>dat</i> that	<i>ik al heel de dag vermoeid</i> I already entire the day tired	<i>ben.</i> am
		‘...that I have been feeling tired throughout the entire day.’		
(38) Contaminating construction: [ADJECTIVE + COPULA]	Ø	<i>dat</i> that	<i>ik al heel de dag blij</i> I already entire the day happy	<i>ben.</i> am
		‘...that I have been happy throughout the entire day.’		

The same kind of potential for contamination arises with *worden*, which can be used as an auxiliary for non-perfective passives as in (39) in Table 4, or as a change-of-state copula as in (40) in Table 4. Again, when the participle is equivocal with regard to its categorial status as either verbal or adjectival, ambiguity arises, as in (41).

**Table 4.** A target instance of the verbal clusters with the auxiliary *worden* ‘become’ and the contaminating construction [ADJECTIVE + COPULA]

	Prefield	1st verbal pole	Midfield	2nd verbal pole
(39) Target instance of the verbal clusters:	Ø	<i>dat</i> that	<i>de situatie door hem</i> the situation by him	<i>verziekt wordt.</i> sickened is
		‘...that the situation is being screwed up by him.’		
(40) Contaminating cxn: [ADJECTIVE + COPULA]	Ø	<i>dat</i> that	<i>de situatie steeds meer verziekt</i> the situation increasingly more ugly	<i>wordt.</i> becomes
		‘...that the situation is increasingly becoming more ugly.’		

(41) *Ik denk dat de situatie verziekt wordt.*

I think that the situation sickened becomes

‘I think the situation will become ugly.’ / ‘I think the situation is being screwed up.’

Interestingly, the word order variation that we find in verbal clusters, where the green order [PARTICIPLE + AUX] semi-freely alternates with the red word order [AUX + PARTICIPLE], does not occur with adjectives: here only the green word order [ADJECTIVE + COPULA] is possible (Haeseryn et al. 1997:1364–1365; Klooster 2001:340–341). The reason is quite simple: adjectives do not have access to the verbal poles and cannot be placed behind the 2nd verbal pole. So example (40) does not allow any permutation of the adjective *verziekt* ‘ugly’ and the copula *wordt* ‘becomes’, just like example (38) does not allow a permutation of the adjective *blij* ‘happy’ and the copula *ben* ‘am’.

This is fertile ground for constructional contamination. We expect to see three things:

- i. There will be a tendency for participles in unequivocally verbal contexts to occur more often in the green order [PARTICIPLE + AUX] with the auxiliaries *zijn* and *worden* to the extent that the participle at issue is more often used as an adjective elsewhere in the corpus. Let us emphasize here that we are *not* talking about cases where there is true ambiguity, as in (41), but about cases where there are contextual cues to enforce a verbal reading of the [PARTICIPLE + *zijn/worden*]. The reason for this tendency would be that in such constructions, exemplified in (36) and (39), the [VERBAL PARTICIPLE + AUX] string is contaminated by the superficially similar [ADJECTIVAL PARTICIPLE + COPULA] string. This is first degree constructional contamination.
- ii. There will be a tendency for participles in unequivocally verbal contexts to occur more often in the green order [PARTICIPLE + AUX] with the auxiliary *hebben* to the extent that the participle at issue is more often used as an adjective elsewhere in the corpus, but probably to a lesser extent than in (i). The reason is that in such combinations, there may be an influence of the adjectival predilection of the participle, even though the participle in the concrete case at hand is unequivocally verbal because it occurs with the auxiliary *hebben* (which hardly combines with predicative adjectives), and even if there is no superficially resembling bigram [ADJECTIVE + COPULA] that directly interferes.<sup>15</sup> This is second degree constructional contamination.
- iii. Adjectives will occasionally violate the grammaticality constraints described in Dutch reference grammars (Haeseryn et al. 1997:1364–4365; Klooster 2001:340–341) by appearing in the red word order [COPULA + ADJECTIVE] if they superficially resemble verbal participles. The reason is constructional contamination in the reverse direction (see Pijpops & Van de Velde 2016:545, Fn. 4 for this effect in partitive genitives).

15. Disregarded here are examples like *Ik denk dat hij z'n haar gekortwiekt heeft* (literally: ‘I think that he his hair short-clipped has/wears’), which do show the intended ambiguity (De Sutter 2005:201; Coussé 2011:619–620) but are vanishingly rare.

To test whether these expectations are indeed borne out in text corpora, we turned to an existing dataset, used in De Sutter (2005: Chapter 5). The data consist of 2,390 observations of two-part clusters with the auxiliaries *hebben*, *zijn*, and *worden* taken from supra-regional quality newspaper material (*De Standaard*) in the CONDIV-corpus (Grondelaers et al. 2000).

As De Sutter was interested in, among other things, the impact of the semantics of the participle on the choice for the red [AUX + PARTICIPLE] vs. green order [PARTICIPLE + AUX], he set out to disambiguate the examples in which clear disambiguating cues were lacking. This involved using an elaborate algorithm, part of which took into account the predilection of the participle to occur as an adjective or as a verb in general. This was exactly the information we needed as well. Unfortunately (for us), De Sutter provided this information only for verb lemmas that occurred in combination with the auxiliaries *zijn* or *worden*, as the *hebben* cases were non-ambiguous. Our goal, by contrast, is not to disambiguate the equivocal instances but, rather, to assess a distributional skew that may be caused by constructional contamination among the *non-ambiguous* cases. This necessitated recalculating the ‘adjectiveness’ of each unequivocal verbal participle. This was done in the following way: for each participle, we calculated the ratio of the number of times it was tagged as an adjective over the total number of occurrences in the Corpus of Spoken Dutch (Oostdijk et al. 2002).<sup>16</sup> This yielded a numeric variable ADJECTIVENESS.

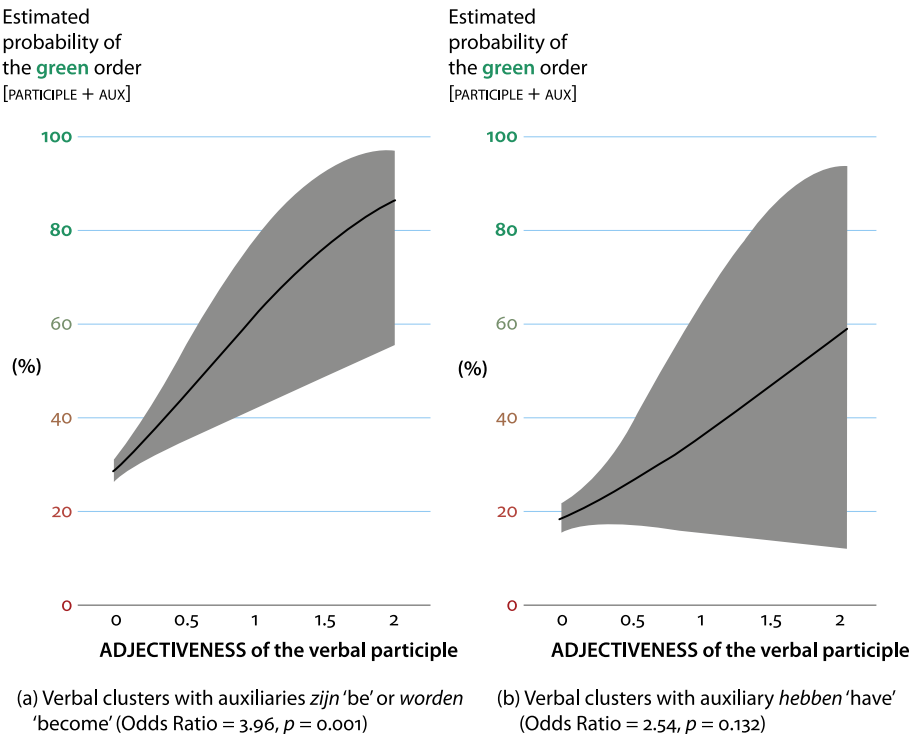
In order to test the expectation in (i), we isolated all observations from De Sutter’s dataset with *zijn* and *worden* as auxiliaries. From the resulting dataset, we excluded all the instances where the participle might be an adjective. This was done by relying on De Sutter’s careful categorization of the participle, which allowed us to discard observations with pseudo-participles, participles with lexical specialization, observations where contextual cues point to an adjectival interpretation, and non-classifiable cases. In other words: only decidedly verbal observations were retained. This yielded a dataset of 1,440 observations. Of this total

---

16. We did this twice, once including the inflected form in attributive contexts and once without the inflected form. Both can be argued for. Not including the inflected form could possibly lead to an underestimation of the number of adjectival occurrences but it has the advantage that the participles used in the calculation are formally identical to the invariably uninflected verbal participle. The two measures are reassuringly highly correlated (Pearson correlation 0.99,  $p < 0.001$ ), so in actual practice, it does not really matter which of the two measures is used. This calculation of ADJECTIVENESS does mean that we are dependent on the POS-tagging scheme of the Corpus of Spoken Dutch. We believe this is justifiable because (i) we agree with its definition of what constitutes an adjective, which is fairly conservative (Van Eynde 2004: 16–22), (ii) it is an objective and easily replicable operationalization, and (iii) we currently see no other operationalization available that is more reliable.

number, 1,005 cases were in the green order [PARTICIPLE + AUX] and 435 cases were in the red order [AUX + PARTICIPLE].

We then fitted a logistic regression that predicted the word order by the ADJECTIVENESS of the verbal participle. ADJECTIVENESS indeed turns out to be significantly associated with word order in unambiguously verbal clusters used with the auxiliaries *zijn* and *worden*, as can also be seen in the effect plot in Figure 1a.<sup>17</sup>



**Figure 1.** Effect plots of ADJECTIVENESS. The higher the ADJECTIVENESS of the participle, the higher the probability of the green order [PARTICIPLE + AUX] in unambiguously verbal clusters, indicating constructional contamination. As predicted, this effect is stronger in (1a) than in (1b), where it does not reach significance.

17. We applied an angular transformation (i.e. the arcsine of the square root) of the ADJECTIVENESS variable. 31 observations were ignored because the ADJECTIVENESS value could not be calculated due to the verbs' absence in the Corpus of Spoken Dutch. All data presented in this paper were analyzed using R (R Core Team 2014), employing the R-packages *dplyr* (Wickham & Francois 2015), *rms* (Harrell 2013), *lme4* (Bates et al. 2013), *effects* (Fox et al. 2016), and *party* (Hothorn et al. 2006).

Testing expectation (ii) involved looking at all observations from De Sutter's dataset with the auxiliary *hebben* ( $n=664$ ).<sup>18</sup> The distribution of the two word order variants is 538 red orders [AUX + PARTICIPLE] vs. 126 green orders [PARTICIPLE + AUX]. We again fitted a regression model to predict the alternation based on the ADJECTIVENESS of the verbal participle. The corresponding effect plot can be inspected in Figure 1b. Here, we see that ADJECTIVENESS also appears to have a positive effect on the probability of a green order [PARTICIPLE + AUX] but the effect size is more moderate and the predictor does not reach significance in the logistic regression. Note that this is not surprising, as we had predicted to find a more moderate effect for the *hebben* cases, which constitute a case of second degree constructional contamination (see (ii) above).

As for expectation (iii), we do indeed observe occasional violations of the supposedly ungrammatical red order with adjectives, as expected. De Sutter (2005:200) gives an observation attested in newspaper material:

- (42) (...) *dat de man al geruime tijd is vermist.*  
           that the man already extensive time is missing  
           '...that the man is missing for a long time already.'

Here, the adverbial *al geruime tijd* strongly suggests an adjectival reading of *vermist* but it occurs in the red order. This is an effect of constructional contamination, as the adjectival participle *vermist* superficially resembles a verbal participle.<sup>19</sup> Other examples can be found easily. The SoNaR corpus contains the following examples of red order with the pseudo-participle *opgewassen* 'able to cope with' and with the adjectival use of *geslaagd* 'successful'.

- (43) *Zij vrezen dat de OR-leden niet zijn opgewassen tegen de*  
       they fear that the OR-members not are able.to.cope with the  
       *onderhandelaars.*  
       negotiators  
       'They fear that the OR-members are no match for the negotiators.'
- (44) *Boonen zelf wist dat hij tegen Petacchi niet was opgewassen.*  
       Boonen himself knew that he with Petacchi not was able.to.cope  
       'Boonen knew himself that he was no match for Petacchi.'

18. For the logistic regression, nine observations were ignored because the ADJECTIVENESS value could not be calculated due to the verbs' absence in the Corpus of Spoken Dutch.

19. In fact, there is no verb *vermissen* that would correspond to *vermist*, attested in present-day Dutch (den Boon & Geeraerts 2005).

- (45) *Wel weet ik dat ik niet **ben opgewassen** tegen zulke grote hoeveelheden onrecht.*  
 PARTICLE know I that I not am able.to.cope with such large quantities  
 injustice  
 ‘I know all too well that I am not up to dealing with such huge amounts of injustice.’
- (46) (...) *dat ze misschien wel **zijn geslaagd** in hun opzet.*  
 that they maybe PARTICLE are succeeded in their plan  
 ‘...that they may have been successful in their plan.’
- (47) *Zijn trainer Arsène Wenger stelt na de huldiging dat het seizoen is **geslaagd**.*  
 his trainer Arsène Wenger asserts after the celebration that the season is  
 successful  
 ‘After the celebration, his trainer Arsène Wenger asserts that the season has been successful.’

Taken together, the three contexts (i–iii) offer support for the effect of constructional contamination. To be sure, the dataset we used does not yield significant results for context (ii), i.e. the *hebben* cases, but the tendency is in the right direction and has a notable effect size. The fact that the effect size was smaller than in context (i), i.e. the *zijn/worden* cases, and the fact that the signal was weaker (higher p-value) in (ii), are actually in line with our expectations, as context (ii) is second degree constructional contamination, whose effect should be less outspoken than first degree constructional contamination (see section 2).

### 3.4 Weak and strong preterites

In this subsection, we present our fourth case of constructional contamination. This one deals with preterite morphology.

In Dutch, like in other Germanic languages, the preterite can be formed in two ways: either a dental suffix is added to the stem of the verb, as in *speelde* ~ *speel* ‘played’ ~ ‘play’, or the vowel of the verb changes, as in *zwom* ~ *zwem* ‘swam’ ~ ‘swim’.<sup>20</sup> The first strategy is called the weak inflection, the second the strong inflection. Often, however, verbs that used to belong to the strong inflection, appear with weak forms. Researchers have proposed many factors that play

20. This two-way distinction ignores reduplication forms (Jasanoff 2007; van Coetsem 1990; von Mengden 2011) and verbs which have both a vowel change and a dental suffix, like *dacht* ~ *denk* ‘thought’ ~ ‘think’.

a role in this ‘weakening’.<sup>21</sup> Two factors, for instance, that are well known to influence this change are the frequency and the ablaut class of the verbs. Lieberman et al. (2007), Carroll et al. (2012), and De Smet & Van de Velde (forthc.), respectively, show for English, German, and Dutch that verbs with a higher frequency tend to weaken less, while Carroll et al. (2012) and Knooihuizen & Strik (2014) observe that verbs belonging to particular ablaut classes are better at resisting the weakening than verbs belonging to certain other classes. Pijpops et al. (2015) present an agent-based simulation explaining why this might be the case.

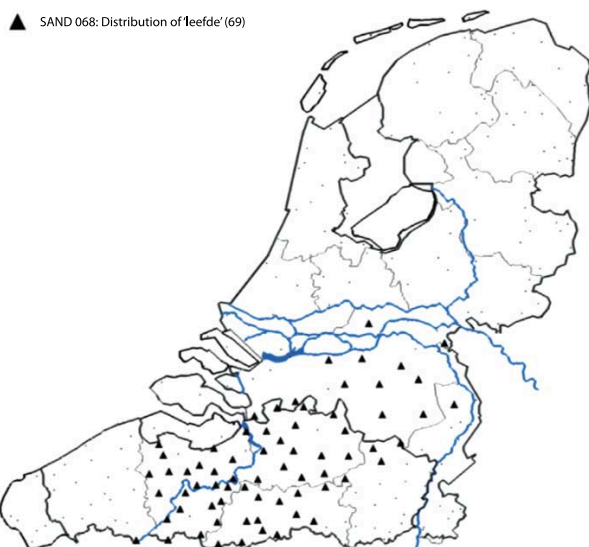
Vosters (2012:242) mentions another possible factor, which he calls “an unusual case of analogy”. This “unusual case of analogy” actually qualifies as a case of constructional contamination. In informal spoken Dutch, there exists a construction that is formally identical to the weak preterite, as in (48), where the second singular subject pronoun is realized as an enclitic (*-de* or *-te*) to the verb.

- (48) Contaminating construction: clitic realization of the present tense, 2nd person singular  
*vandaag graaf-de een put.* (Vosters 2012:242)  
 today dig-2SG.PRS a hole  
 ‘You will dig a hole today.’

As can be seen in Figure 2, these constructions are typical of the central region of Belgium, notably the provinces of Antwerp, Flemish-Brabant, and East-Flanders – though they do sporadically appear north of the state border as well. It is exactly in this central Belgian region that Vosters observed the highest percentage of weak verbs in his data. This suggests that the presence of the clitic expressions might contribute to the weakening of the verbs in this area.

The case study that we bring to bear on this distribution consists of two parts. First, we will use corpus data to try to replicate the results obtained by Vosters (2012), who used a production task. For the second part of this case study, we will explore whether the effects of constructional contamination are also lexically specific. That is, we will investigate whether verbs that are more often realized with a clitic subject tend to weaken more than verbs that are only rarely realized with a clitic subject, comparing for instance *loopte* ‘you ran’ and *’slinkte* ‘you lessened’, as suggested by Vosters (2012:244, Fn. 16). Our hypotheses are the following:

21. We prefer the term ‘weakening’ over the perhaps better known term ‘regularization’ because the latter implies (i) that the weak inflection is fully regular, which it is not (cf. the weak forms *sent* and *built* in English, and *kocht* ‘bought’ in Dutch), and (ii) that the strong inflection is void of regularity, which it is neither, at least in Dutch (cf. Pijpops et al. 2015:79–81).



**Figure 2.** Dialectal spread of enclitic subject pronouns in inverted syntactic position (Vosters 2012: 243, based on data from the Syntactic Atlas of the Dutch Dialects (SAND, Barbiers et al. 2006))

- i. Weak preterites will be more prevalent in the regions known for their enclitic realization of the subject pronoun, compared to the other Dutch-speaking regions of the Low Countries.
- ii. Verbs that are more often realized with an enclitic subject tend to weaken more than verbs that are less often realized with an enclitic subject.

For the first part, we needed a Dutch corpus with regional information. We used a corpus of Dutch tweets from 2012 that contains geolocation information.<sup>22</sup> In this corpus, we looked for both weak and strong attestations of the strong verbs mentioned in Haeseryn et al. (1997: 432–441). We only included simplex verbs and their corresponding complex forms with non-separable particles. We did not include the most frequent verbs (*worden* ‘become’, *zitten* ‘sit’, *komen* ‘come’, *staan* ‘stand’, *gaan* ‘go’, *vinden* ‘find’, *zijn* ‘be’, *zien* ‘see’) as they are unlikely to weaken anyway, nor did we include verbs that are only strong in jocular language use (e.g. *foof* ~ *fuif* ‘party’) or verbs in which the choice of inflection is dependent upon its meaning (e.g. *pleegde* ‘committed’ vs. *placht* ‘used to’ ~ *pleeg*). Because our corpus is not morphologically annotated, we had to exclude verbs of which the weak preterite plural would be the same as the present or infinitive (e.g. *bidden* ~ *bid*

22. The corpus was compiled for another study by Tom Ruette, to whom we want to express our gratitude for sharing it.



‘pray’). We also had to exclude verbs of which the strong form was homonymous with frequent nouns (e.g. *school* ~ *schuil* ‘school’ or ‘hide’, *stal* ~ *steel* ‘steal’ or ‘stable’ etc.). Lastly, we excluded all verbs that were found to exhibit no variation in our corpus and invariably appeared in the weak or strong inflection. We manually went through all attestations ending in *-de* or *-te* in order to separate the forms with enclisis from actual weak preterites. This resulted in a dataset containing 3,490 strong forms and 151 weak forms.<sup>23</sup>

Based on this dataset, we built a mixed-effects regression model with the inflection of the occurrence as a binary response variable and the weak inflection as the success level. Regional provenance of the tweet and frequency of the verb were added as fixed effects. The variable *REGION* bears our primary interest. It distinguishes between, on the one hand, central Belgium, defined as one of the central Dutch-speaking provinces of Antwerp, Flemish-Brabant, and East-Flanders, and the remaining Dutch-speaking regions in the Low Countries on the other hand. The *FREQUENCY* of the verb was added as a control variable. It is counted as the natural logarithm of the division between the verb’s number of occurrences in the Corpus of Spoken Dutch (Oostdijk et al. 2002) and the total number of verbs in the same corpus. Finally, the lemma of the verb was added as a random effect and the model was fitted to the data. The details of this regression model can be found in Table 5.

The model confirms our hypothesis: we find a significant effect ( $p = 0.031$ ) for region and an Odds Ratio of 0.395 indicating that the weak inflection is indeed less prevalent in the regions outside of central Belgium.<sup>24</sup> This corroborates the findings of Vosters (2012).

For the second part of the study, where we are interested in the correlation between a verb’s proclivity for the weak preterite and its proclivity to occur with an enclitic pronoun, we evidently needed to know which verbs appear more often with an enclitic pronoun. Because the previous investigation yielded no more than 45 attestations of the enclitic expressions, this required additional data and therefore a larger corpus. As the enclitic variant appears almost exclusively in informal Belgian Dutch, we opted for the informal components of the SoNaR corpus, i.e. the tweets, chats, discussion fora, and text messages, and further restricted

23. Our dataset contains the following verbs: *bevelen* ‘order’, *dragen* ‘carry’, *drijven* ‘float’, *duiken* ‘dive’, *ervaren* ‘experience’, *jagen* ‘hunt’, *klagen* ‘complain’, *krimpen* ‘shrink’, *lezen* ‘read’, *raden* ‘guess’, *roepen* ‘shout’, *schelden* ‘scold’, *stoten* ‘stub’, *vangen* ‘catch’, *vliegen* ‘fly’, *vreten* ‘devour’, *waaien* ‘blow’, and *zuigen* ‘suck’ (and their complex verbs with non-separable particles).

24. As can be seen in Table 6, *FREQUENCY* also has a highly significant effect on the realization of the preterite: verbs with a lower frequency had a higher probability of weak preterites, again confirming the observations of, among others, Lieberman et al. (2007), Carroll et al. (2012), and De Smet & Van de Velde (forthc.).

**Table 5.** Numeric output of the mixed-effects regression model with inflection as binary response variable, the weak inflection as the success level, and verb lemma as random factor

Explanatory variable	Level	Estimate	Confidence interval		P-value
			2.5%	97.5%	
REGION	intercept	-18.809	-28.741	-11.275	<0.001
	enclisis region	reference level			
	non-enclisis region	-0.930	-1.820	-0.018	0.031
RELATIVE FREQUENCY (log-transformed)		-2.032	-3.164	-1.174	<0.001

us to the Belgian material. We extracted all occurrences of the same verbs as in the first part of this case study, though this time we did not have to exclude as many homonyms, because SoNaR is lemmatized. However, since the tagging is not always reliable, we still had to exclude *vroeg* ~ *vraag* ‘asked’ or ‘early’, *at* ~ *eten* ‘ate’ or the symbol ‘@’ written out, and *wrong* ~ *wringen* ‘wrung’ or the English borrowing (or codeswitch) *wrong*. Other possibly ambiguous verbs that caused less noise were dealt with manually. Again, only verbs that show variation were included. This dataset contained 12,097 attestations of strong preterites and 220 attestations of weak preterites.<sup>25</sup> We also went through the weak preterites manually to sort out the enclitic forms from the real preterites, this time dividing the number of enclitic forms for each verb by the total frequency of each verb in order to compose a new variable, ENCLITIC PROCLIVITY. The total FREQUENCY of the verbs was calculated based on the same informal subcorpora of SoNaR that we used to get our data. Again, we ran a mixed-effect logistic regression with the realization of the preterite as the outcome variable, the log-transformed relative FREQUENCY and the ENCLITIC PROCLIVITY as predictors, and the verb lemma as random effect.<sup>26</sup> We observe a trend in the expected direction, yet the tendency is

25. The verbs in this dataset are: *bieden* ‘offer’, *blazen* ‘blow’, *blijken* ‘prove’, *dragen* ‘carry’, *drinken* ‘drink’, *duiken* ‘dive’, *dwingen* ‘force’, *graven* ‘dig’, *hangen* ‘hang’, *helpen* ‘help’, *jagen* ‘hunt’, *kiezen* ‘choose’, *klagen* ‘complain’, *klinken* ‘sound’, *knijpen* ‘squeeze’, *laten* ‘let’, *liggen* ‘lie’, *lijden* ‘suffer’, *lopen* ‘run’, *raden* ‘guess’, *roepen* ‘shout’, *ruiken* ‘smell’, *schelden* ‘scold’, *schenden* ‘violate’, *schrijven* ‘write’, *schuilen* ‘hide’, *slapen* ‘sleep’, *slinken* ‘diminish’, *sluiten* ‘close’, *springen* ‘jump’, *spugen* ‘spit’, *steken* ‘stab’, *stelen* ‘steal’, *sterven* ‘die’, *stijgen* ‘rise’, *stinken* ‘stink’, *strijken* ‘iron’, *trekken* ‘pull’, *vangen* ‘catch’, *varen* ‘sail’, *vechten* ‘fight’, *waaien* ‘blow’, *wegen* ‘weigh’, *werpen* ‘throw’, *zuigen* ‘suck’, *zuipen* ‘booze’, *zweren* ‘swear’, *zwijgen* ‘keep silent’ and their complex verbs with non-separable particles.

26. An angular transformation (the arcsine of the square root) was applied to the ENCLITIC PROCLIVITY variable.

not significant ( $p=0.729$ ), which can be seen in Table 6. The reason for this may lie in our highly skewed dataset, containing only 220 weak forms versus 12,097 strong forms, or in the fact that the contaminating enclitic construction is not frequent enough in our data (19 verbs with 2,354 observations in total had an ENCLITIC PROCLIVITY of 0).<sup>27</sup>

**Table 6.** Numeric output of the mixed effects regression model with inflection as the binary outcome variable, the weak inflection as the success level, and verb lemma as random factor

Explanatory variable	Level	Estimate	Confidence interval		P-value
			2.5%	97.5%	
	intercept	-12.787	-15.843	-9.827	<0.001
ENCLITIC PROCLIVITY (angular-transformed)		3.555	-17.105	24.415	0.729
RELATIVE FREQUENCY (log-transformed)		-1.323	-1.727	-0.925	<0.001

#### 4. Conclusions

In this article, we have presented three new case studies of constructional contamination. The first concerned the competition between the bare infinitive (without *te* 'to') and the long infinitive (with *te* 'to'). It was found that constructional

27. If we take out the verb lemma as a random effect (as a mixed-model with a highly skewed distribution in the levels of the outcome variable will have a hard time reliably estimating the coefficients), the p-value decreases ( $p=0.28$ ). If, on top of that, we remove all instances with an ENCLITIC PROCLIVITY of zero, the p-value decreases even more and we obtain a significant effect of enclitic proclivity (with  $p=0.012$ ). Removing all the instances with ENCLITIC PROCLIVITY=0 could be motivated as this value conflates verbs that are categorically excluded from occurring with an enclitic 2SG pronoun with verbs that just happen not to be attested with such an enclitic pronoun in our dataset. For instance, *slinken* 'decrease' hardly ever occurs in the 2SG, as it mostly takes inanimate subjects. Its zero-value on ENCLITIC PROCLIVITY is more severe than the zero-value of, say, *spugen* 'spit', which is semantically plausible in 2SG but which is not attested in this context in our dataset. As a consequence, there might be much unexplained variance within the zero-group. These are post-hoc considerations however, and we do not want to engage in p-value hacking. Given the rather extreme difference in the distribution of the two outcome variable levels, we might consider conditional inference trees (see Tagliamonte & Baayen 2012) as a non-parametric alternative for regression. This yields significance for two splits in the (angular-transformed) ENCLITIC PROCLIVITY ( $p<0.001$  for (angular-transformed) ENCLITIC PROCLIVITY > 0.056 and  $p=0.007$  for (angular-transformed) ENCLITIC PROCLIVITY > 0.018), adding further support to the constructional contamination hypothesis.

contamination may be powerful enough to bend the borders of grammaticality (see also Van de Velde 2015 and Van de Velde & Pijpops 2018). We observed that the bare infinitive was most frequent in contexts affected by first degree constructional contamination, i.e. those exhibiting direct superficial similarity to the contaminating construction; it was less frequent in contexts affected by second degree constructional contamination, i.e. those exhibiting similarity to first degree contexts; and it was vanishingly rare in any other context.

The second case study investigated word order in verbal clusters. It was shown that the word order of adjectives affected the preferences of superficially similar verbal participles. First degree constructional contamination was again shown to generate stronger effects than second degree contamination. In addition, we found bidirectional constructional contamination (as predicted in Pijpops & Van de Velde 2016:545, Fn.4), as adjectives were demonstrated to appear in word orders that are normally grammatically restricted to verbal participles, under the condition that these adjectives superficially resemble verbs. This again indicates that constructional contamination is capable of defying the limits of grammaticality.

The third case study dealt with the preterite. Here, we succeeded in replicating the results of the production task by Vosters (2012) using corpus data. We found a significant effect in the expected direction. Dialects of Dutch that feature the contaminating enclitic construction were also shown to favor the affected weak forms. Digging deeper, we also tested whether the effect of constructional contamination was lexically specific (see Vosters 2012:244, Fn. 16). We did find a weak tendency in the expected direction, yet it was not significant. Perhaps the contaminating enclitic construction was not frequent enough to generate a clearly measurable, lexically specific effect on an alternation that is already stringently determined by other factors, such as token frequency.

Including the partitive genitive construction reported on in Pijpops & Van de Velde (2016), we now have four case studies of constructional contamination: two from morphology and two from syntax. All four yield positive results – albeit to varying degrees. We therefore believe it is safe to regard constructional contamination as a fairly wide-spread phenomenon, affecting both morphology and syntax. This does not mean that it is a sure-fire factor in alternation studies, though. For one, its effect may be eclipsed by more dominant factors, as we suspect may have been the case in the follow-up investigation of the case study on preterite formation. For another, in case studies where ambiguity-avoidance is a pressing matter, it could be possible that the superficial formal overlap in condition B is explicitly avoided, thereby blocking constructional contamination. More research is therefore needed to determine the precise contexts that favor constructional contamination.

Such research can take at least three forms. The first is the form of corpus studies, such as the ones presented in this paper. Grammatical alternations as defined in the tradition of alternation research (Gries 2003; Grondelaers 2000; Heller 2018; Pijpops & Speelman 2017; Szmrecsanyi 2017) all fulfill condition A, i.e. a formal alternation between two interchangeable variants, and several of these probably also meet condition B, i.e. a superficial resemblance between some of their instances and another construction (see section 2). Many possible cases of contaminating constructions may still fly under the radar. A good, although work-intensive, method of tracking down such possibly contaminating constructions is the manual checking of corpus instances, as argued in Pijpops & Van de Velde (2016:577). Some other, automated methods of identifying contaminating instances are already explored in Pijpops & Van de Velde (2016) but further research on these would still be welcome. Finally, corpus research on constructional contamination may be extended to other languages, as well as to lexicology and phonology.

The second form that future research on constructional contamination may take is psycholinguistic experiments. Constructional contamination is theorized to be the side-effect of a specific routine of language processing, i.e. exemplar chunking. As such, experiments seem the ideal methodology to study it. Still, the effect is the result of long-term exposure to a contaminating construction, which may be difficult to achieve within the limited timespan of an experiment. However, this is only a practical difficulty which can no doubt be overcome in a clever research design. For examples on experiments on superficial parsing, see Dąbrowska (2014) and the overviews in Ferreira & Patson (2007) and Dąbrowska (2015).

The third line of research is that of computer simulation. In this paper, we have assumed that the processing strategies described in Ferreira & Patson (2007), Bybee (2013), and Dąbrowska (2014) do indeed bring about constructional contamination as a side effect. For now, this assumption is solely based on verbal speculation. If these processing strategies were to be computationally implemented, we could test whether constructional contamination does indeed emerge. Ideally, such an implementation would enable us to set up a feedback loop between empirical research and computer simulation, such that corpus studies and experiments could reveal how constructional contamination behaves in reality and simulation could inform us what exact assumptions we need to implement in order to explain this behavior. Such a feedback loop has, for instance, been set up for research on the acquisition of the English past tense (see e.g. Bybee & Slobin 1982; Rumelhart & McClelland 1986; Pinker & Prince 1988; Marcus et al. 1995; Taatgen

& Anderson 2002).<sup>28</sup> Some examples of work in this direction are Data-Oriented Parsing (DOP; Scha et al. 1999; Bod 2006, 2009) and Fluid Construction Grammar (FCG; Steels 2011, 2017; van Trijp et al. 2012). For more examples, see the overview in Beuls & van Trijp (2016).

We would like to end this article by summarizing the relevance of constructional contamination for three branches of linguistic inquiry. First, concerning research on language processing, the most plausible explanation of the effect of constructional contaminating is that language users can indeed work with large, unanalyzed chunks of language in memory, as proposed by Bybee (2010:2013) and Dąbrowska (2014, 2015) and the ‘good-enough’ approach to language processing advocated in Ferreira et al. (2002) and Ferreira & Patson (2007). Second, concerning alternation studies, constructional contamination can constitute an additional factor to explain the choice between two linguistic variants, extending the typical dichotomy between language-internal (semantic and syntactic) variables and lectal variables (age, gender, SES, genre, register, etc.) in Labovian variational linguistics, with processing-related factors. Some work has already been done here (e.g. Szmrecsanyi 2005 and Szmrecsanyi et al. 2016 on ‘ $\beta$ -persistence’, and Claes & Johnson, forthcoming, on markedness of coding, statistical preemption, and structural priming), and constructional contamination can be added to this list. Third, concerning Construction Grammar theory, constructional contamination highlights the importance of horizontal resemblance links in the constructicon, as argued for by, among others, Van de Velde (2014) and Norde & Trousdale (2016). In particular, it shows how mere superficial formal resemblance may overcome structural distinctions between constructions and cause interference between them. Constructional contamination shows how these interferences can be accounted for through such horizontal links (cf. De Smet et al. 2018: 11–12). To summarize, constructional contamination offers a rare window on language processing, a promising explanation of language variation, and a concrete test-case for language theory.

## Acknowledgements

This research benefited from a BOF-grant at the University of Leuven and funding from the Research Foundation Flanders (FWO). We want to thank Gert De Sutter for generously sharing

---

28. The gist of the idea behind such simulations is to implement the relevant acquisition or processing strategy without explicitly programming in their hypothesized results (e.g. U-shaped learning in research on the acquisition of the English past tense, or constructional contamination in our case) and then to observe whether these effects do indeed emerge as a by-product.

his dataset of two-verb clusters collected for his PhD (De Sutter 2005), as well as Tom Ruetten for giving us access to his Twitter corpus. We also owe thanks to two anonymous reviewers for highly valued comments and feedback.

## References

- Augustinus, L., & Van Eynde, F. (2017). A usage-based typology of Dutch and German IPP verbs. *Leuvense Bijdragen: Tijdschrift Voor Germaanse Filologie*. Nijhoff.
- Barbiers, S., Bennis, H., De Vogelaer, G., Devos, M., & van der Ham, M. (2006). *Syntactic atlas of the Dutch dialects. Vol. 1: Pronouns, Agreement and Dependencies*. Amsterdam: Amsterdam university press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.4*. Retrieved from <http://cran.r-project.org/package=lme4>
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M., Croft, W., Ellis, N., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(1), 1–26. <https://doi.org/10.1111/j.1467-9922.2009.00533.x>
- Beuls, K., & van Trijp, R. (2016). Computational construction grammar and constructional change. *Belgian Journal of Linguistics*, 30, 1–13. <https://doi.org/10.1075/bjl.30.01beu>
- Bloem, J. (2016). Lexical preferences in Dutch verbal cluster ordering. In K. Bellamy, E. Karvovskaya, M. Kohlberger, & G. Saad (Eds.), *ConSOLE XXIII: Proceedings of the 23rd Conference of the Student Organization of Linguistics in Europe* (pp. 70–93). Leiden: Leiden University Centre for Linguistics.
- Bloem, J., Versloot, A., & Weerman, F. (2014). Applying automatically parsed corpora to the study of language variation. In J. Tsujii & J. Hajic (Eds.), *Proceedings of COLING 2014: the 25th International Conference on Computational Linguistics: technical papers* (pp. 1974–1984). Dublin: Dublin City University and Association for Computational Linguistics
- Bloem, J., Versloot, A., & Weerman, F. (2015). An agent-based model of Germanic verbal cluster word order change. In *Computational Linguistics in the Netherlands (CLIN)*. February 6, Antwerp.
- Bloem, J., Versloot, A., & Weerman, F. (2017). Verbal cluster order and processing complexity. *Language Sciences*, 60, 94–119. <https://doi.org/10.1016/j.langsci.2016.10.009>
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *Linguistic Review*, 23(3), 291–320. <https://doi.org/10.1515/TLR.2006.012>
- Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science*, 33(5), 752–793. <https://doi.org/10.1111/j.1551-6709.2009.01031.x>
- Bolinger, D. (1980). Wanna and the gradience of auxiliaries. In G. Brettschneider & C. Lehmann (Eds.), *Wege zur Universalienforschung: sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler* (pp. 292–299). Tübingen: Gunter Narr.
- Booij, G. (2010). Construction morphology. *Language and Linguistics Compass*, 4(7), 543–555. <https://doi.org/10.1111/j.1749-818X.2010.00213.x>
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

- Broekhuis, H., & Strang, A. (1996). De partitieve genitiefconstructie [The partitive genitive construction]. *Nederlandse Taalkunde*, 1(3), 221–238.
- Bybee, J. (2010). *Language, usage, and cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49–69). Oxford: Oxford University Press.
- Bybee, J., & Slobin, D. (1982). Rules and schemas in the development and use of the English past tense. *Language*, 58(2), 265–289. <https://doi.org/10.1353/lan.1982.0021>
- Carroll, R., Svare, R., & Salmons, J. (2012). Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics*, 2(2), 153–172. <https://doi.org/10.1075/jhl.2.2.01car>
- Claes, J., & Johnson, D. E. (Forthcoming). Cognitive linguistics and the predictability of effects: Agreement in English and Spanish existentials.
- Colleman, T. (2009). Verb disposition in argument structure alternations: A corpus study of the dative alternation in Dutch. *Language Sciences*, 31(5), 593–611. <https://doi.org/10.1016/j.langsci.2008.01.001>
- Coupé, G. (2015). *Syntactic extension. The historical development of Dutch verb clusters*. Leiden: LOT.
- Coussé, E. (2008). Motivaties voor volgordevariatie. Een diachrone studie van werkwoordsvolgorde in het Nederlands. Dissertation, University of Ghent.
- Coussé, E. (2011). On ambiguous past participles in Dutch. *Linguistics*, 49(3), 611–634. <https://doi.org/10.1515/ling.2011.019>
- Dąbrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics*, 25(4), 617–653. <https://doi.org/10.1515/cog-2014-0057>
- Dąbrowska, E. (2015). Language in the mind and in the community. In J. Daems, E. Zenner, K. Heylen, & D. Speelman (Eds.), *Change of paradigms – new paradoxes. Recontextualizing language and linguistics* (pp. 221–235). Berlin: De Gruyter.
- De Smet, H., D'hoedt, F., Fonteyn, L., & Van Goethem, K. (2018). The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics*, 29(2), 197–234. <https://doi.org/10.1515/cog-2016-0025>
- De Smet, I. & F. Van de Velde (Forthcoming, 2019). Reassessing the evolution of West Germanic preterite inflection. *Diachronica* 36(2).
- De Sutter, G. (2005). Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen. Dissertation, University of Leuven.
- den Besten, H., & Edmonson, J. (1983). The verbal complex in continental West-Germanic. In W. Abraham (Ed.), *On the formal syntax of the Westgermania* (pp. 155–216). Amsterdam: John Benjamins. <https://doi.org/10.1075/la.3.05bes>
- den Boon, T., & Geeraerts, D. (Eds.). (2005). *Van Dale Groot woordenboek van de Nederlandse taal* (14th ed). Antwerpen/Utrecht: Van Dale Lexicography.
- Diessel, H. (2015). Usage-based construction grammar. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 296–322). Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110292022-015>
- Ferreira, F., Bailey, K., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>



- Ferreira, F., & Patson, N. (2007). The “good enough” approach to language comprehension. *Language and Linguistics Compass*, 1, 71–83. <https://doi.org/10.1111/j.1749-818X.2007.00007.x>
- Fox, J., Weisberg, S., Friendly, M., Hong, J., Andersen, R., Firth, D., & Taylor, S. (2016). *Effect displays for linear, generalized linear, and other models. R package version 3.2.*
- Geeraerts, D. (2010). Recontextualizing grammar: Underlying trends in thirty years of Cognitive Linguistics. In E. Tabakowska, M. Choinski, & L. Wiraszka (Eds.), *Cognitive linguistics in action: From theory to application and back* (pp. 71–102). Berlin/New York: De Gruyter Mouton.
- Gibson, E. A. F. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Dissertation, Carnegie Mellon University.
- Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement.* New York: Continuum.
- Grondelaers, S. (2000). De distributie van niet-anaforisch er buiten de eerste zinsplaats: sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal. Dissertation, University of Leuven.
- Grondelaers, S., Deygers, K., Van Aken, H., Van den Heede, V., & Speelman, D. (2000). Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpus of written Dutch]. *Nederlandse Taalkunde*, 5(4), 356–363. Retrieved from <http://neon.niederlandistik.fu-berlin.de/static/digitaal/digitaal-11.html>
- Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. (1997). *Algemene Nederlandse Spraakkunst* [General Dutch grammar]. Groningen: Nijhoff.
- Harrell, F. (2013). rms: Regression modeling strategies. R package version 4.0-0. Retrieved from <http://cran.r-project.org/package=rms>
- Heine, B. (1993). *Auxiliaries: cognitive forces and grammaticalization.* New York: Oxford University press.
- Heller, B. (2018). Stability and fluidity in syntactic variation world-wide. The genitive alternation across varieties of English. Dissertation, University of Leuven.
- Hoeksema, J. (1998). Adjectivale inflectie op -s: geen geval van transpositie [Adjectival inflection on -s: not a case of transposition]. In E. Hoekstra & C. Smits (Eds.), *Morfologiedagen 1996* [Morphology Days 1996] (pp. 46–72). Amsterdam: P. J. Meertens-Instituut.
- Hopper, P. (1987). Emergent grammar. *Berkeley Linguistic Society*, 13, 139–157. <https://doi.org/10.3765/bls.v13i0.1834>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Jasanoff, J. H. (2007). From reduplication to ablaut: The class VII strong verbs of Northwest Germanic. *Historische Sprachforschung / Historical Linguistics*, 120, 241–284. <https://doi.org/10.13109/hisp.2007.120.1.241>
- Kester, E.-P. (1996). *The nature of adjectival inflection.* LEd, Utrecht. Utrecht: LEd
- Klooster, W. (2001). *Grammatica van het hedendaags Nederlands: een volledig overzicht* [Grammar of contemporary Dutch: A complete overview]. Den Haag: Sdu.
- Knooihuizen, R., & Strik, O. (2014). Relative productivity potential of Dutch verbal inflection patterns. *Folia Linguistica Historica*, 35(1), 173–200.
- Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1178–1183). Stresa: Cognitive Science Society.

- Konieczny, L., Müller, D., Hachmann, W., Schwarzkopf, S., & Wolfer, S. (2009). Local syntactic coherence interpretation. Evidence from a visual world study. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1133–1138). Austin: Cognitive Science Society.
- König, E., & Gast, V. (2009). *Understanding English-German contrasts* (2nd edn.). Berlin: Erich Schmidt.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
- Lemmens, M. (2005). Aspectual posture verb constructions in Dutch. *Journal of Germanic Linguistics*, 17(3), 183–217. <https://doi.org/10.1017/S1470542705000073>
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163), 713–716. <https://doi.org/10.1038/nature06137>
- Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29(3), 189–256. <https://doi.org/10.1006/cogp.1995.1015>
- Norde, M., & Trousdale, G. (2016). Exaptation from the perspective of construction morphology. In M. Norde & F. Van de Velde (Eds.), *Exaptation and language change* (pp. 163–195). Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.336.06nor>
- Oostdijk, N., Goedertier, W., Van Eynde, F., Boves, L., Martens, J.-P., Moortgat, M., & Baayen, H. (2002). Experiences from the Spoken Dutch corpus project. In *Proceedings of the third international conference on language resources and evaluation (LREC)* (pp. 340–347). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2002/>
- Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The Construction of a 500-million-word reference corpus of contemporary written Dutch. In P. Spyns & J. Odiijk (Eds.), *Essential speech and language technology for Dutch, theory and applications of natural language processing* (pp. 219–247). Heidelberg: Springer. [https://doi.org/10.1007/978-3-642-30910-6\\_13](https://doi.org/10.1007/978-3-642-30910-6_13)
- Pauwels, A. (1953). *De plaats van hulpwerkwoord verleden deelwoord en infinitief in de Nederlandse bijzin*. Leuven: Symons.
- Phillips, C. (1996). Order and structure. Dissertation, Massachusetts Institute of Technology.
- Pijpops, D., Beuls, K., & Van de Velde, F. (2015). The rise of the verbal weak inflection in Germanic. An agent-based model. *Computational Linguistics in the Netherlands Journal*, 5, 81–102.
- Pijpops, D., & Speelman, D. (2017). Alternating argument constructions of Dutch psychological verbs. A theory-driven corpus investigation. *Folia Linguistica*, 51(1), 207–251. <https://doi.org/10.1515/flin-2017-0006>
- Pijpops, D., & Van de Velde, F. (2015). Ethnolect speakers and Dutch partitive adjectival inflection. A corpus analysis. *Taal En Tongval*, 67(2), 343–371. <https://doi.org/10.5117/TET2015.2.PIJP>
- Pijpops, D., & Van de Velde, F. (2016). Constructional contamination: How does it work and how do we measure it? *Folia Linguistica*, 50(2), 543–581. <https://doi.org/10.1515/flin-2016-0020>
- Pijpops, D., & Van de Velde, F. (2018). A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion. *Corpus linguistics and linguistic theory*, 14(1), 99–131. <https://doi.org/10.1515/cllt-2013-0027>

- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1), 73–193. [https://doi.org/10.1016/0010-0277\(88\)90032-7](https://doi.org/10.1016/0010-0277(88)90032-7)
- R Core Team. (2014). *R: A language and environment for statistical computing*. R foundation for statistical computing. Vienna. Retrieved from <http://www.r-project.org/>
- Röthlisberger, M., Grafmiller, J., & Szmrecsanyi, B. (2017). Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics*, 28(4), 673–710. <https://doi.org/10.1515/cog-2016-0051>
- Rumelhart, D., & McClelland, J. (1986). On learning the past tense of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 216–271). Cambridge: MIT Press.
- Scha, R., Bod, R., & Sima'an, K. (1999). A memory-based model of syntactic analysis: Data-oriented parsing. *Journal Of Experimental & Theoretical Artificial Intelligence*, 11(3), 409–440. <https://doi.org/10.1080/095281399146481>
- Schmid, T. (2005). *Infinitival syntax: Infinitivus pro participio as a repair strategy*. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.79>
- Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, & H.-P. Schwefel (Eds.), *Proceedings of PPSN VI: Lecture notes in computer science* (pp. 17–26). Berlin: Springer.
- Steels, L. (2011). *Design patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins. <https://doi.org/10.1075/cal.11>
- Steels, L. (2017). Basics of fluid construction grammar. *Constructions and Frames*, 9(2), 178–225. <https://doi.org/10.1075/cf.00002.ste>
- Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1(1), 113–150. <https://doi.org/10.1515/cllt.2005.1.1.113>
- Szmrecsanyi, B. (2017). Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique*, 62(4), 685–701. Retrieved from <https://muse.jhu.edu/article/678232>
- Szmrecsanyi, B., Biber, D., Egbert, J., & Franco, K. (2016). Toward more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change*, 28(1), 1–29. <https://doi.org/10.1017/S0954394515000198>
- Taatgen, N., & Anderson, J. (2002). Why do children learn to say “broke”? A model of learning the past tense without feedback. *Cognition*, 86, 123–155. [https://doi.org/10.1016/S0010-0277\(02\)00176-2](https://doi.org/10.1016/S0010-0277(02)00176-2)
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355–370. <https://doi.org/10.1016/j.jml.2004.01.001>
- Tagliamonte, S., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. <https://doi.org/10.1017/S0954394512000129>
- van Bart, P., Kerstens, J., & Sturm, A. (1998). *Grammatica van het Nederlands. Een inleiding* [Grammar of Dutch. An introduction]. Amsterdam: Amsterdam University Press. <https://doi.org/10.5117/9789053562819>
- van Coetsem, F. (1990). *Ablaut and reduplication in the Germanic verb*. Heidelberg: Winter.

- Van de Velde, F. (2014). Degeneracy: The maintenance of constructional networks. In R. Boogaart, T. Coleman, & G. Rutten (Eds.), *Extending the scope of construction grammar*, Vol. 1 (pp. 141–179). Berlin: Mouton de Gruyter.
- Van de Velde, F. (2015). Schijnbare syntactische feniksen [Apparent syntactic phoenixes]. *Nederlandse Taalkunde*, 20(1), 69–107. <https://doi.org/10.5117/NEDTAA2015.1.VELD>
- Van de Velde, F. (2017). Understanding grammar at the community level requires a diachronic perspective. Evidence from four case studies. *Nederlandse Taalkunde*, 22(1), 47–74. <https://doi.org/10.5117/NEDTAA2017.1.VELD>
- Van de Velde, F. & D. Pijpops. 2018. Grensoverschrijdend syntactisch gedrag [Cross-border syntactic behavior]. In T. Coleman, J. De Caluwe, V. De Tier, A.-S. Ghyselen, L. Triest, R. Vandenberghé & U. Vogl (Eds.), *Woorden om te bewaren. Huldeboek voor Jacques Van Keymeulen* [Words to preserve. Articles in honor of Jacques Van Keymeulen] (pp. 433–449). Ghent: UGent, Department of Linguistics, research group of Dutch.
- van der Horst, J. (2008). *Geschiedenis van de Nederlandse syntaxis* [History of Dutch syntax]. Leuven: Universitaire Pers Leuven.
- Van Eynde, F. (2004). Part of speech tagging en lemmatisering van het corpus gesproken nederlands. Retrieved from <http://www.hum.uu.nl/medewerkers/p.monachesi/papers/vaneynde.pdf>
- van Noord, G. (2006). At last parsing is now operational. In P. Mertens, C. Fairon, A. Dister, & P. Watrin (Eds.), *TALN 2006. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles* (pp. 20–42). Louvain-la-Neuve: Cental.
- van Trijp, R. (2008). Analogy and multi-level selection in the formation of a case grammar. A case study in Fluid Construction Grammar. Dissertation, University of Antwerp.
- van Trijp, R., Steels, L., Beuls, K., & Wellens, P. (2012). Fluid construction grammar: The new kid on the block. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 63–68). Avignon: ACL.
- von Mengden, F. (2011). Ablaut or transfixation? On the Old English strong verbs. In R. Bauer & U. Krischke (Eds.), *More than words: English lexicography and lexicology past and present. Essays presented to Hans Sauer on the occasion of his 65th birthday – Part I.* (pp. 123–139). Frankfurt am Main: Lang.
- Vosters, R. (2012). Geolinguistic data and the past tense debate. Linguistic and extralinguistic aspects of Dutch verb regularization. In G. De Vogelaer & G. Seiler (Eds.), *The dialect laboratory. Dialects as a testing ground for theories of language change* (pp. 227–248). Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/slcs.128.10vos>
- Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal Commitment theory goes east. *Journal of Psycholinguistic Research*, 22(3), 339–364. <https://doi.org/10.1007/BF01068016>
- Wickham, H., & Romain, F. (2015). *dplyr: A grammar of data manipulation*. R package version 0.4.3. <https://CRAN.R-project.org/package=dplyr>
- Zifonun, G., Hoffmann, L., & Strecker, B. (1997). *Grammatik der deutschen Sprache* [Grammar of the German language]. Berlin: de Gruyter.
- Zwart, J.-W. (2011). *The syntax of Dutch*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511977763>

## Address for correspondence

Dirk Pijpops  
KU Leuven, FWO  
Blijde Inkomststraat 21, box 3308  
3000 Leuven  
Belgium  
dirk.pijpops@kuleuven.be

## Co-author information

Isabeau De Smet  
KU Leuven, FWO  
isabeau.desmet@kuleuven.be

Freek Van de Velde  
KU Leuven  
freek.vandavelde@kuleuven.be