Peter Dirix/Liesbeth Augustinus/Frank Van Eynde (Leuven)

IPP in Afrikaans: a corpus-based investigation and a comparison with Dutch and German

Abstract: In the West-Germanic languages we expect an auxiliary of the perfect to select a past participle. In a subset of these languages, however, some verbs select an infinitive instead, i.e. in constructions known as *infinitivus pro participio* (IPP). The phenomenon is well-studied with regard to Dutch and German, but for Afrikaans an extensive study based on empirical data is still lacking. In order to fill this void, the present paper uses a corpus study to identify the verbs which – obligatorily or optionally – take the IPP form in Afrikaans. Verb classes showing the IPP effect in Afrikaans, Dutch and German are compared, and crosslinguistic similarities and differences are identified. The result is a corpus-based typology of IPP verbs in the three languages in question.

Zusammenfassung: In den westgermanischen Sprachen wäre eigentlich zu erwarten, dass Perfekt-Hilfsverben immer ein Partizip Perfekt selektieren. In einer Untergruppe dieser Sprachen selektieren einige Verben jedoch einen Infinitiv, den so genannten *infinitivus pro participio* (IPP). Während dieses Phänomen hinsichtlich des Niederländischen und Deutschen bereits eingehend erforscht worden ist, fehlt zum Afrikaans bisher eine umfangreichere, empirisch fundierte Studie. Um diesem Mangel abzuhelpen, werden in dem vorliegenden Beitrag mittels einer Korpusuntersuchung diejenigen Verben identifiziert, die im Afrikaans – obligatorisch oder optional – in der IPP-Form auftreten. Wir vergleichen die Verbklassen, die auf Afrikaans, Niederländisch und Deutsch den IPP-Effekt zeigen, und stellen Ähnlichkeiten und Unterschiede zwischen den Sprachen fest. Das Ergebnis ist eine korpusbasierte Typologie von IPP-Verben in den drei betroffenen Sprachen.

1 Introduction

In the West-Germanic languages the auxiliary of the perfect combines with a past participle (PSP), as illustrated in (1–6) for German (DE), Dutch (NL), Afrikaans (AF), English (EN), West Frisian (FY), and Yiddish (YI).¹

- (1) DE Wir haben nichts *gesehen*.
- (2) NL Wij hebben niets *gezien*.
- (3) AF Ons het niks *gesien* nie.
- (4) EN We have not *seen* anything.
- (5) FY Wy hawwe neat *sjoen*.
- (6) YI Mir hobn gornisht *gezen*.

If the participle combines with an infinitival complement, though, it may and sometimes must take the form of an infinitive, as illustrated in (7–9).

- (7) DE Ich habe Johann einen Roman schreiben *sehen/gesehen*.
- (8) NL Ik heb Johan een roman *zien*/**gezien* schrijven.
- (9) AF Ek het Johan 'n roman *sien/gesien* skryf.
'I saw John write a novel.'

The phenomenon is known as *infinitivus pro participio* (IPP).² It occurs in German, Dutch and Afrikaans, but not in English, Frisian or Yiddish (Wurmbrand 2004; Schmid 2005: 138). It is an interesting topic for comparative research, because it shows a large degree of variation. Notice, for instance, that the Dutch *zien* in (8) must take the IPP form, while its counterparts in German and Afrikaans may take either the usual participial form or the IPP form.

While German and Dutch IPP have been studied extensively, Afrikaans IPP has received far less attention. The first objective of this study therefore is to identify the verbs which – obligatorily or optionally – take the IPP form in Afrikaans. For this purpose we make use of two corpora. The second objective is to compare the results with those for Dutch and German. In answering this question, we use the typology of IPP verbs that has been proposed in Augustinus/Van Eynde (2017).

¹ We thank the audience of the Germanic Sandwich conference (March 2017, Münster), the anonymous reviewers and Alexander Hurst for their comments.

² Other terms are *Ersatzinfinitiv* and *substitute infinitive*, see Den Besten/Edmondson (1983).

The paper starts with a survey of the literature on IPP in Afrikaans in section 2. In section 3 we present the corpus study. Section 4 describes the typology of Augustinus/Van Eynde (2017) and compares it to other lists and classifications for German and Dutch IPP verbs. In addition, an extension to the typology is proposed, making use of the data from the corpus study. Section 5 summarizes our findings.

2 IPP in Afrikaans

This section provides a brief survey of the literature on IPP in Afrikaans. To pave the way we first present some facts about the verbal paradigm (2.1). Then we discuss the constructions in which the IPP effect occurs (2.2). A separate section is devoted to the modal verbs (2.3).

2.1 The verbal paradigm

Afrikaans verbs do not show much inflectional variation. The finite forms do not show any variation for person or number and the present/past distinction is – for most verbs – not morphologically marked (De Vos 2005). Instead the past is marked by the perfect auxiliary *het* ‘have’ in combination with the past participle, as illustrated for the verb *bly* ‘to stay’.

ek bly ‘I stay’	ek het gebly ‘I stayed’
jy/u ³ bly ‘you (sing.) stay’	jy/u het gebly ‘you (sing.) stayed’
hy/sy/dit bly ‘he/she/it stays’	hy/sy/dit het gebly ‘he/she/it stayed’
ons bly ‘we stay’	ons het gebly ‘we stayed’
julle/u bly ‘you (plur.) stay’	julle/u het gebly ‘you (plur.) stayed’
hulle bly ‘they stay’	hulle het gebly ‘they stayed’

The past participle canonically has the prefix *ge-*, but this prefix is omitted if the verb already has an unstressed prefix, such as *be-*, *ge-*, or *ver-*, as shown in *jy het begin* ‘you started’.⁴ The only verbs with a morphologically marked past tense, henceforth called preterite, are the copula and most of the modals.

³ *Jy* is the informal form. The polite form *u* is hardly ever used in present-day Afrikaans.

⁴ A form like *gebegin* is sometimes observed in colloquial language.

<i>ek is</i> ‘I am’	<i>ek was</i> ‘I was’
<i>ek kan</i> ‘I can’	<i>ek kon</i> ‘I could’
<i>ek wil</i> ‘I want’	<i>ek wou</i> ‘I wanted’
<i>ek sal</i> ‘I will’	<i>ek sou</i> ‘I would’
<i>ek moet</i> ‘I must (pres.)’	<i>ek moes</i> ‘I must (past)’
<i>ek mag</i> ‘I may’	<i>ek mog</i> ‘I might’ ⁵

The infinitive has the same form as the simple present tense, except in the case of the copula *wees* ‘be’ and the main verb *hê* ‘have’.⁶ They contrast respectively with *is* and *het* (Donaldson 1993).

The passive is formed by the combination of a passive auxiliary and a past participle. The auxiliary is *word* ‘become’ for the present and *is* ‘be’ for the past. Compare the active *sy slaan hom* ‘she hits him’ and *sy het hom geslaan* ‘she hit him’ with the passive counterparts *hy word deur haar geslaan* ‘he is hit by her’ and *hy is deur haar geslaan* ‘he was hit by her’. In colloquial speech one also hears *was* (instead of *is*), but this is considered an Anglicism, being a double past.

2.2 IPP

In Afrikaans, IPP mainly occurs in the double infinitive construction (2.2.1) and in pseudo-coordination (2.2.2). A special case is IPP in passive constructions (2.2.3).

2.2.1 Double infinitive construction

As pointed out in the introduction, the perfect auxiliary canonically combines with a past participle, as in (3) and (10), but if the participle selects an infinitival complement, it may take the IPP form instead, as in (9) and (11).

- (10) Hy het stil gebly/*bly.
 he has quiet remain.PSP/remain.INF
 ‘He remained quiet.’

⁵ *Mog* ‘might’ is archaic and does not appear anymore in contemporary Afrikaans. The verb *hê* ‘have’ used to have a preterite form (*had*), but this is now replaced by *het gehad*. The modals *hoef* ‘need to’ and *behoort* ‘ought to’ do not have a preterite form. Their past counterparts are formed with the auxiliary *het*.

⁶ For the homophonous auxiliary the infinitive is identical to the simple present, i.e. *het*.

- (11) Hy het gebly/bly praat.
 he has remain.PSP/remain.INF talk.INF
 'He kept talking.'

Ponelis (1979), De Vos (2001) and Zwart (2007) claim that the choice for the IPP form is optional, but Donaldson observes that it is often preferred in the contexts where it is allowed and that the use of the past participle in (10) is typical of colloquial Afrikaans (Donaldson 1993: 225f.). If the IPP form is chosen, one gets a sequence of two infinitives, hence the term *double infinitive construction*.

A list of verbs that may take the IPP form is provided in Ponelis (1979). It includes the aspectual verbs *begin* 'begin', *bly* 'stay', *aanhou* 'continue', *ophou* 'stop', *gaan* 'go', *kom* 'come' and *loop* 'walk', the causative *laat* 'let', and the control verbs *help* 'help', *leer* 'learn, teach' and *probeer* 'try'. They all combine with a bare infinitive, except for *begin*, which can also have an infinitive introduced by *te* 'to'.

Van Schoor (1983) contains a longer list, which also includes *kry* 'get', *durf* 'dare', the causative *maak* 'make', the evidential modals *skyn* 'seem' and *blyk* 'turn out', and the perception verbs *hoor* 'hear', *sien* 'see', *voel* 'feel' and *ruik* 'smell'. According to Van Schoor, IPP is obligatory for all of these verbs, except for *aanhou* and *ophou*.

Table 1: IPP versus PSP (Robbers 1997)

	IPP		PSP	
<i>aanhou</i> 'continue'	0	(0%)	0	(0%)
<i>begin</i> 'begin'	0	(0%)	0	(0%)
<i>bly</i> 'remain'	17	(94.44%)	1	(5.56%)
<i>gaan</i> 'go'	71	(100%)	0	(0%)
<i>help</i> 'help'	1	(33.33%)	2	(66.67%)
<i>hoor</i> 'hear'	2	(66.67%)	1	(33.33%)
<i>laat</i> 'let'	49	(98%)	1	(2%)
<i>leer</i> 'learn/teach'	5	(100%)	0	(0%)
<i>sien</i> 'see'	16	(94.12%)	1	(5.88%)
<i>voel</i> 'feel'	1	(100%)	0	(0%)
Total	162	(96.43%)	6	(3.57%)

Robbers (1997) is the first to provide quantitative data. For a subset of the IPP verbs she counted the frequency of IPP and non-IPP (i.e. PSP) occurrences in a 128K-word corpus, consisting of (parts of) three novels.⁷ The results are presented in Table 1.

This survey clearly shows that IPP is preferred when it is possible, confirming the claim in Donaldson (1993). The only exception is *help*, but notice that there are only three occurrences of this verb in the data set.

Note that Conradie (2006) claims there is no such thing as IPP in Afrikaans.

- (12) Hy het 'n liedjie *laat* /*gelaat* sing.
 he has a song let.INF /let.PSP sing.INF
 'He made them sing a song.'

In a construction like (12), he analyses *laat sing* as a single verb, and considers (*ge*)*laat sing* as a single past participle with optional *ge*- prefixation. He does not give any arguments for this analysis, though. According to Ponelis (1993), also cited in Robbers (1997), a small number of these combinations of a linking verb and a main verb have become lexicalized, e.g. *gaan haal* 'fetch', *laat blyk* 'indicate', *laat geld* 'exercise authority', *laat kom* 'summon', *laat spaander* 'get going' and *laat staan* 'leave'. Most combinations, however, are not lexicalized as the verb clustering is a productive process and there is no specialization of meaning. The present tense version is "Hy *laat* 'n liedjie *sing*", making it hard to claim *laat sing* is a single verb.

2.2.2 Pseudo-coordination

Another construction in which IPP forms occur concerns a serialization pattern with the conjunction *en* 'and'. It is used to express the continuous or progressive aspect, as in (13).

- (13) Ons staan en luister.
 we stand.PRES and listen.INF
 'We are listening.'

Since the conjunction does not have its usual coordinating function here, the construction is known as *pseudo-coordination*. It also exists in English (e.g. *he sits*

⁷ Robbers' corpus study is based on novels of Van Heerden (1987, first 100 pages), Van Niekerk (1994, first 100 pages) and Brink (1995, first 160 pages), but she does not mention the titles herself.

and reads), but not in Dutch or German. If *staan* is combined with the perfect auxiliary, it optionally takes the IPP form. The verb after *en*, by contrast, must take the infinitival form, as shown in (14).

- (14) Ons het gestaan/staan en luister/*geluister.
 we have stand.PSP/stand.INF and listen.INF/listen.PSP
 ‘We were listening.’

Both options are considered standard Afrikaans by Ponelis (1979: 241–245), Donaldson (1993), Zwart (2007), and Verdoolaege/Van Keymeulen (2010). Robbers (1997) points out that *en* may be omitted, as in (15–16).

- (15) Sy staan ril effe.
 she stands tremble.INF momentarily
 ‘She is trembling for a minute.’
- (16) Sy het gestaan/staan ril effe.
 she has stand.PSP/stand.INF tremble momentarily
 ‘She was trembling for a minute.’

Notice that the variant with the IPP form in (16) has the same form as the double infinitive construction.

Beside *staan* ‘stand’, the verbs which occur in this construction are *lê* ‘lie’, *loop* ‘walk’, and *sit* ‘sit’.⁸ Notice that this use of *loop* is different from the one that it has in the double infinitive construction. Robbers (1997) provides some quantitative data, see Table 2.

Table 2: IPP versus PSP in pseudo-coordination (Robbers 1997)

	IPP		PSP	
<i>lê</i> ‘lie’	2	(50%)	2	(50%)
<i>loop</i> ‘walk’	0	(0%)	2	(100%)
<i>sit</i> ‘sit’	2	(16.67%)	10	(83.33%)
<i>staan</i> ‘stand’	3	(15%)	17	(85%)
Total	7	(22.58%)	31	(77.42%)

⁸ Ponelis (1979) points out that the infinitival marker *te* ‘to’ appears instead of *en* in a very limited number of cases. Such constructions are archaic.

In contrast to what we observed for the DIC, we see a preference for the non-IPP combination.

2.2.3 Passive IPP constructions

De Vos (2001) reports that some of the IPP verbs, esp. *laat* ‘let’, tend to passivize fairly productively. Examples are (17–18).

- (17) Hierdie huis is deur my oom laat /gelaat bou.
 this house is by my uncle let.PSP /let.INF build
 ‘My uncle had this house built.’
- (18) ’n stuk hout wat daarheen laat sak word
 a piece wood that to-there let.INF lower.INF is
 ‘a piece of wood that is lowered to that place’

Robbers (1997: 61–64) also provides examples with *kom* ‘come’, *ophou* ‘stop’ and *begin* ‘begin’. If the embedded infinitive is a transitive verb, such as *bou* ‘build’, its object is identified with the subject of the passive auxiliary, as in (17). If the embedded infinitive is intransitive, it is its subject which is identified with the subject of the passive auxiliary, as in (18).⁹

According to Ponelis (1979), complex passives are not acceptable in pseudo-coordination. De Vos (2001) also points out that, although speaker judgements vary, it is difficult to passivize constructions with pseudo-coordination. Robbers (1997), however, had two informants who accept the construction, and Breed (2012) cites a few examples from the internet, such as (19).

- (19) ? Die appel word deur hom gesit en eet.
 The apple is by him sit.PSP and eat.INF
 ‘He sits and eats the apple’

Notice that *sit* has the PSP form. Breed’s examples all contain transitive verbs, such as *eet* ‘eat’.

⁹ Constructions like (17) are impossible in Dutch and German. However, some speakers of German allow the *remote passive construction*, e.g. *Weil der Wagen oft zu reparieren versucht wurde*. ‘Because many attempts were made to repair the car.’ (Müller 2002: 136). In those constructions, the object of the embedded verb is realized as the subject of the selecting verb (in this case *versuchen* ‘try’). In German, these constructions do not show the IPP effect.

2.2.4 Summing up

The literature on Afrikaans discusses a limited set of verbs that figure in IPP constructions. They concern the double infinitive construction, pseudo-coordination and their passive counterparts. The literature is inconclusive about the issue of which verbs belong to the set (each author has his/her own set) and whether IPP is optional or obligatory. With a more extensive corpus study we aim to shed some more light on this issue, see section 3.

2.3 Modal verbs

The modal verbs are special in two respects. First, they are – together with the copula – the only verbs with a morphologically marked past tense, see 2.1. Second, they lack a past participle form. From this it follows that the past tense counterpart of a sentence with a modal verb is formed in another way than in the case of the other verbs. As an example, let us take the sentence *Jan kan hard werk* ‘John can work hard’. According to Robbers (1997) it has no less than five possible past tense counterparts. The most common one is (20), in which the modal has the preterite form.¹⁰

(20) Jan kon hard werk. [pret₁ – inf₂]

The alternatives all contain the auxiliary of the perfect *het*.¹¹

(21)	Jan het hard kon werk	[aux ₁ – pret ₂ – inf ₃]
(22)	Jan het hard kan werk	[aux ₁ – pres ₂ – inf ₃]
(23)	Jan kon hard gewerk het	[pret ₁ – psp ₃ – aux ₂]
(24)	Jan kan hard gewerk het	[pres ₁ – psp ₃ – aux ₂]

In (21) the auxiliary is combined with the preterite form *kon*, which in turn selects the infinitive *werk*. Robbers calls it archaic. In (22) the auxiliary selects the present

¹⁰ In the added annotation *aux* stands for the auxiliary *het* ‘have’, *pres* for the present tense of the modal, *pret* for the preterite form of the modal, *inf* for the infinitival form of the main verb, and *psp* for the participial form of the main verb. The subscripts indicate the order of selection, with the hierarchically highest verb being 1.

¹¹ Ponelis (1979) calls the construction in (21) *preteritive assimilation* and the one in (23) *preteritive replacement*.

form *kan* which selects *werk*. Robbers doubts whether it is well-formed. In (23) and (24) the main verb takes the form of the past participle which is selected by the auxiliary *het*; the modal appears in its preterite form in (23) and in its simple present form in (24). Ponelis (1993: 412) suggests that the preterite form of the modal in (21) crossed over from finite to infinitival territory.

We find the same possibilities in subordinate clauses, but with another word order.

(25)	... dat Jan hard kon werk	[pret ₁ – inf ₂]
(26)	... dat Jan hard kon werk het	[pret ₂ – inf ₃ – aux ₁]
(27)	... dat Jan hard kan werk het	[pres ₂ – inf ₃ – aux ₁]
(28)	... dat Jan hard kon gewerk het	[pret ₁ – psp ₃ – aux ₂]
(29)	... dat Jan hard kan gewerk het	[pres ₁ – psp ₃ – aux ₂]

For verbs that have identical forms for the infinitive and the past participle, such as *gebeur* ‘happen’ and *verkie* ‘prefer’, the [pret₂–inf₃–aux₁] sequence in (26) is not distinct from the [pret₁–psp₃–aux₂] sequence in (28), and the [pres₂–inf₃–aux₁] sequence in (27) is not distinct from the [pres₁–psp₃–aux₂] sequence in (29).

An obvious question is whether there are any semantic differences between the five alternatives. Stell (2011: 162–165) points out that the patterns in (21) and (22) are only used with declarative (indicative) senses, whereas the canonical pattern in (20) and the patterns with the past participle in (23) and (24) are also used with conditional and hypothetical senses, including the irrealis.

3 Corpus study

For the corpus study we have used two corpora: the *Taalkommissie* corpus and the Afrikaans section of Wikipedia.

The *Taalkommissie* corpus is a 54 million-word corpus of written Afrikaans, developed by the Centre for Text Technology (CTeX 2011) of North-West University’s Potchefstroom campus under the auspices of the language committee of the *Suid-Afrikaanse Akademie vir Wetenskap en Kuns* (South African Academy of Science and Arts). It covers a wide selection of genres including newspapers and magazine articles, textbooks, law and government texts, Bible texts, and literature excerpts. Unfortunately, the version of the corpus at our disposal does not contain any metadata, so it is not possible to track the origin of the sentences or to compare the different registers. The corpus was automatically tokenized, annotated with part-of-speech (PoS) tags and integrated in a search tool as described

in Augustinus/Dirix (2013). The tag set consists of 139 different tags, mainly based on morphosyntactic features (Pilon 2005). The author of the tagger claims an accuracy of 85.87% on a small data set, which is rather low compared to state-of-the-art PoS taggers for well-resourced languages. Schlünz (2010) reports an accuracy of 94.64%, but with a tag set reduced to only 17 different tags, mainly just part-of-speech. We decided to use the extended tag set, as the tags for the different verb forms are important to us.

Wikipedia¹² is the world's best known online encyclopaedia, available in about 280 languages. Its Afrikaans section contained 40,820 articles with 13.2 million words at the moment of download (July 2016). After extracting the articles with WikiExtractor.py,¹³ we applied the same tokenization and tagging procedure as we did for the *Taalkommissie* corpus.

The following sections describe the results of the corpus investigation of both corpora. Section 3.1 presents the double infinitive constructions, section 3.2 pseudo-coordination, section 3.3 passive constructions, and section 3.4 constructions with modal verbs. Section 3.5 compares the results for the two corpora. Section 3.6 concludes.

3.1 Double infinitive constructions

In order to collect double infinitive constructions and the corresponding constructions with a past participle, we searched for the following constructions in the *Taalkommissie* and Wikipedia corpora:¹⁴

- Constructions in which the auxiliary *het* 'have' is immediately followed by two verbs (main clauses),
- Constructions in which the auxiliary *het* 'have' is immediately preceded by two verbs (subordinate clauses),
- Constructions in which there is another word between *het* 'have' and the two other verbs (main clauses).

¹² www.wikipedia.org (last accessed: 12-6-2016).

¹³ Downloadable from <https://github.com/attardi/wikiextractor> (last accessed: 12-6-2016).

¹⁴ In Augustinus/Dirix (2013) a more limited corpus study of IPP in the *Taalkommissie* corpus was done. Constructions with modal verbs and verbs selecting a *te* infinitive were not considered in that study. The methodology for extracting double infinitive constructions and constructions with pseudo-coordination is similar to the methodology used for this study.

It is obviously possible that more than one word occurs between *het* and the two other verbs, see example (9) in the introduction, but we limited our research to constructions with no more than one word between *het* and the verbal group. Searching for constructions with more words in between resulted in data that were too noisy to manually classify.¹⁵

The corpus searches resulted in more than 13,000 matches, which were manually checked and categorized. After filtering out the false positives, we retained 7,713 hits, of which 7,632 contain an IPP verb.

The results of the corpus investigation are presented in Table 3. The first column describes the verb types. The categorization is taken from Augustinus/Van Eynde (2017).¹⁶ The classification is based on the distinction between subject- and object-oriented VP selectors. In IPP constructions with a subject-oriented verb, the unexpressed subject of the infinitival complement is identified with the subject of the IPP, whereas in IPP constructions with an object-oriented verb the unexpressed subject is identified with the (in)direct object of the IPP. An example of a subject-oriented verb is *try* in *He tried to cheat on us*, whereas *made* in *She made him smile* is an instance of an object-oriented verb. The subject-oriented verbs are further divided into aspectual, evidential and subject-control verbs, while the object-oriented verbs are partitioned into causative, perception and benefactive verbs.

The second column of Table 3 mentions the lemma of the verbs. The third and fourth column contain the corpus results from the *Taalkommissie* corpus and Wikipedia respectively. The fifth column contains the row totals, and the sixth column mentions the percentage of hits that contain an IPP construction for the verb under investigation. Within the corpus results (third and fourth column), the first column represents the verb's occurrence as an IPP verb selecting a bare infinitive (IPP), the second column indicates how often it occurs as an IPP verb selecting a *te* infinitive (IPP + *te*), the third column indicates how often the verb occurs in a construction with a past participle (PSP).

¹⁵ Using a treebank would solve this problem, but unfortunately the only treebank at our disposal, AfriBooms (Augustinus et al. 2016; Dirix et al. 2017), turned out to be too small to investigate the IPP phenomenon. Therefore, we decided to use flat corpora, but with a restriction on the length of the constructions under investigation.

¹⁶ In section 4 we will compare the results of this study to the typology of IPP verbs proposed in Augustinus/Van Eynde (2017).

Table 3: Double infinitive constructions and corresponding constructions with a past participle in the Taalkommissie corpus and Wikipedia

Type	Verb	Taalkommissie corpus			Wikipedia			Total	% IPP	
		IPP	IPP + <i>te</i>	PSP	IPP	IPP + <i>te</i>	PSP			
subject-oriented	aspectual	<i>begin</i> ‘begin’	1,461	1	1	713	1	1	2,178	99.91
		<i>gaan</i> ‘go’	853	0	0	200	0	0	1,053	100.00
		<i>kom</i> ‘komen’	645	0	17	147	0	0	809	97.90
		<i>bly</i> ‘continue’	270	0	1	87	0	0	358	99.72
		<i>aanhou</i> ‘continue’	45	0	6	24	0	3	78	88.46
		<i>ophou</i> ‘stop’	16	0	6	10	0	0	32	81.25
	subject control	<i>probeer</i> ‘try’	575	0	1	172	0	0	748	99.87
		<i>durf</i> ‘dare’	35	0	1	4	0	0	40	97.50
		<i>leer</i> ‘learn’	24	0	6	3	0	5	38	71.05
		<i>weet</i> ‘manage’	0	3	0	0	0	0	3	100.00
	evidential	<i>blyk</i> ‘turn out’	0	3	4	0	3	2	12	50.00
object-oriented	causative	<i>laat</i> ‘let’	1,502	0	2	507	0	0	2,011	99.90
		<i>maak</i> ‘make’	1	0	5	1	0	0	7	28.57
	perception verbs	<i>sien</i> ‘see’	130	0	7	9	0	1	147	94.56
		<i>hoor</i> ‘hear’	4	0	0	0	0	0	4	100.00
	benefactive verbs	<i>help</i> ‘help’	111	0	8	70	0	3	192	94.27
		<i>leer</i> ‘teach’	2	0	1	0	0	0	3	66.67
Total		5,674	7	66	1,947	4	15	7,713	98.95	

The results in Table 3 show that there are very few examples without IPP: 7,632 constructions or 98.95% of the matches contain an IPP construction. For 11 of the 17 verbs the percentage of IPP forms is higher than 90% and for three of those it is even 100%. Examples and more detailed discussion are given in 3.1.1 for the subject-oriented and in 3.1.2. for the object-oriented VP selectors.

We also observed there are 84 examples of clusters with three verbs in the cluster and 4 examples with four verbs. All of these have IPP,¹⁷ as in example (30):

¹⁷ This is also true in Dutch for clusters with three verbs or more.

- (30) Sy sou graag meermale wou kom speel het,
 she would happily more-times want.PRET come.INF play.INF have.INF
 maar die opsigter gee haar die krees.
 but the supervisor gives her the creeps
 ‘She would have come over to play more often, but the supervisor gives her the creeps.’

3.1.1 Subject-oriented VP selectors

The set of **aspectual verbs** is the largest category, both with respect to verb types (6) and tokens (4,508). They add up to 58.45% of the 7,713 constructions under investigation. Some examples are given in (31–32).¹⁸

- (31) Hy het *gaan* draf.
 he has go.IPP run
 ‘He went running.’
- (32) Sy het *begin* huil en het my uitgesit omdat ek dronk was.
 she has begin.IPP cry and has me out-thrown because I drunk was
 ‘She started crying and threw me out because I was drunk’

The verbs *begin* ‘begin’, *gaan* ‘go’, *kom* ‘come’, and *bly* ‘continue’ appear as IPP in more than 97% of the cases. The verbs *aanhou* ‘continue’ and *ophou* ‘stop’ have a slightly higher percentage of constructions with a past participle, cf. (33–34).

- (33) Ek het vir hom gesê die kluis staan oop en dis leeg, maar hy
 I have for him said the safe stands open and this-is empty, but he
 het *aangehou* vra.
 has keep.PSP ask.INF
 ‘I said to him the safe was open and empty, but he kept asking.’
- (34) Hy het *opgehou* oefen na sy motorongeluk 20 jaar gelede.
 he has stop.PSP practise after his car-accident 20 years ago
 ‘He stopped practicing 20 years ago after his car accident.’

¹⁸ All examples in this section are taken from the *Taalkommissie* corpus. We consider the use of *begin* in (32) to be an IPP, rather than a PSP.

Note that those verbs are the only verbs with a separable particle encountered in the data set.

The figures in the PSP column of the *Taalkommissie* corpus include several combinations with two participles. For *kom* ‘come’ this is the case for 12 out of the 17 PSP constructions, cf. example (35).

- (35) 'n Vragmotor wat in die teenoorgestelde rigting aangery
a truck which in the opposite direction drive.PSP
gekom het ...
come.PSP has
‘A truck which came from the opposite direction ...’

The literature does not mention this construction.¹⁹ Our consultants consider them ill-formed or very colloquial.²⁰ This might explain why we did not encounter this construction in the Wikipedia corpus.

The **subject control verbs** account for 10.75% of the constructions under investigation. An example is given in (36).

- (36) Hy het *probeer* terugbakei, maar is in 'n stoeiery in die bors
he has try.INF fight-back.INF but is in a struggle in the chest
geskiet ...
shot
‘He tried to fight back, but was shot in the chest during a struggle ...’

In contrast to Dutch *proberen* ‘try’, which is an optional IPP verb, the Afrikaans *probeer* has a high preference for IPP; we have encountered only one construction with a past participle and it clearly contains colloquial language (37).

- (37) oom Gert Wiese het *geprobeer* paai en gesê: “Broer, laat ons
uncle Gert Wiese has try.PSP appease.INF and say.PSP brother let us
tog maar’ie vrede bewaar en da moet ons ôk
PRT but-the.COLL peace keep and then.COLL must we also.COLL
onthou ons is nog altyd innie voorhowe vannie
remember we are still always in-the.COLL forecourts of-the.COLL

¹⁹ We also encountered one instance of a double participle construction for the verb *bly* ‘continue’ and two for the verb *sien* ‘see’ in the *Taalkommissie* corpus.

²⁰ We have asked five native speakers whether they could use such constructions.

tempel.”

temple

‘Uncle Gert Wiese tried to appease him and said: “Brother, let us keep the peace and we must also keep in mind that we are still in the forecourts of the temple.”’

Similar to the aspectual verbs IPP is the preferred construction for Afrikaans subject control verbs. Only in the case of *leer* ‘learn’ have we encountered a substantial percentage (nearly 30%) of constructions with a past participle, as in (38).

- (38) Sy het *geleer* fantaseer, en aanhou fantaseer, deur haar
 she has learn.PSP fantasize.INF, and keep.INF fantasize.INF through her
 skooljare, deur haar studentedae ...
 school-years through her student-days
 ‘She learned to fantasize and kept fantasizing, during her years at school,
 during her days at college ...’

The examples with *geleer* are definitely not colloquial, so *leer* can be considered a truly optional IPP verb in Afrikaans. In Wikipedia, the use of the past participle is the preferred option, but note that the overall frequency of the verb is too small to draw hard conclusions on this matter.

The verb *weet* ‘know’ is one of the few verbs that select a *te* infinitive in IPP constructions, cf. example (39).

- (39) Dis sy ma, het die man *weet* te vertel, wat hom so
 this-is his mother, has the man know.INF to tell.INF, which him so
 genadeloos gedruk het toe hy nog jonk was.
 relentlessly pressured has when he still young was
 ‘The man managed to say that it is his mother who pressured him relentlessly when he was still young.’

Another verb that exclusively selects a *te* infinitive is the **evidential verb** *blyk* ‘turn out’, as used in (40).

- (40) Dit het heel doeltreffend *blyk* te wees en die masjien het die
 this has very efficient seem.INF to be.INF and the machine has the
 eerste in Groot-Brittanje (deur Reynolds) vrygestel en het groot treë
 first in Great-Britain (by Reynolds) released and has great steps
 vorentoe in die fietstegnologie getoon.
 forward in the bicycle-technology showed

‘It seemed very efficient and the machine released the first one in Great Britain (by Reynolds) and showed a big step forward in bicycle technology.’

Blyk is the only verb in this category. It occurs both as IPP and past participle, so IPP is clearly optional for this verb, but note that the overall frequencies of *blyk* are low in both corpora (0.16% of the constructions under investigation). As an IPP verb it only appears in the construction *blyk te wees* ‘appear to be’. Other verbs that would fit into this category are *lyk* ‘seem’ and *skyn* ‘appear’. Those verbs appear in the corpus, but not in combination with *het*.²¹

3.1.2 Object-oriented VP selectors

Within the set of object-oriented VP selectors the **causatives** form the largest category (2018 hits, 26.16%). Most of the constructions are instances of *laat* ‘let’ (41), but we also found a few examples with *maak* ‘make’ in the sense of ‘make someone do something’, as in (42).

- (41) Na 'n dol dag op die filmstel, met 'n regisseur wat die een toneel na
after a crazy day on the film-set, with a director which the one scene after
die ander tot vervelens toe *laat* herhaal het ...
the other until boring to let.INF repeat.INF has
‘After a crazy day on the film set, with a director who made us repeat one
scene after another ad nauseam ...’
- (42) Haal die Haaie die eindstryd, sou dit wees omdat hulle tereg
reach the Sharks the finals, would this be because they rightly
hul veldtog op die basiese hoekstene van die spel. *maak*
their campaign on the basic cornerstones of the game make.INF
staan het.
stand.INF have
‘If the Sharks reach the finals, it would be because they rightly based their
campaign on the fundamentals of the game’

²¹ The Dutch cognates *blijken* ‘turn out’, *schijnen* ‘appear’, and *lijken* ‘seem’ show similar properties: constructions with IPP are hard to find as those verbs are rarely combined with the auxiliary of the perfect (Broekhuis/Corver 2015: 621).

Maak ‘make’ is the only verb in Table 3 for which the IPP hits are outnumbered by the PSP hits (28.57%). Notice that its Dutch cognate *maken* is not used as an IPP at all.

The **perception verbs** *hoor* ‘hear’ and *sien* ‘see’ account for a meagre 1.96% of the constructions. Examples are given in (43) and (44).

- (43) ... sodat sy hulle nooit hoor terugkom het nie.
 ... so-that she them never hear.INF return.INF has not
 ‘... so she never heard them return’
- (44) omdat sy hom nog nooit so sien glimlag het nie.
 because she him yet never so see.INF smile.INF has not
 ‘because she had never seen him smile that way’

Both verbs clearly prefer IPP. For *hoor* we did not find any instances of PSP, and for *sien* just a few, including (45).

- (45) ’n ADT-veiligheidswag wat die voorval gesien gebeur het, het sy
 a ADT-security-guard which the event see.PSP happen.INF has,has his
 noodknoppie gedruk om bystand te ontbied.
 emergency-button pressed for assistance to send-for
 ‘An ADT security guard who witnessed the incident, pushed the emer-
 gency button to call assistance.’

The set of **benefactives** concludes the typology. The 195 hits account for 2.53% of the constructions. Its members *leer* ‘teach’ and *help* ‘help’ both occur as IPP constructions and with a participle:

- (46) Emily Petlo, ’n buurvrou wat Klyn help skoonmaak het, sê die
 Emily Petlo, a neighbour which Klyn help.INF clean.INF has, says the
 water in haar toiletbak staan ook hoog.
 water in her toilet stands also high
 ‘Emily Petlo, a neighbour who helped Klyn to clean, said the water in her
 toilet was also high.’
- (47) Net soos my pa my geleer werk het, so leer ek my kinders.
 just like my father me teach.PSP work.INF has, so teach I my children
 ‘Just as my father taught me how to work, I am teaching my children.’

3.1.3 Summing up

The results in Table 3 show that the aspectual verbs are the prototypical IPP verbs in Afrikaans, followed by causative *laat* ‘let’ and the set of subject control verbs. Evidential *blyk* ‘turn out’, causative *maak* ‘make’ and benefactive *leer* ‘teach’ occur in less than 70% of the hits in the IPP form, but note that the absolute frequencies for those verbs are very low. In general, the double infinitive construction with IPP is by far the preferred construction compared to constructions in which a past participle is selected.

If we compare the verbs in Table 3 to the verbs mentioned in section 2, we see that the results from the corpus study include all verbs mentioned in Ponelis (1979), except for *loop* ‘run’. Van Schoor (1983) also mentions the verbs *kry* ‘get’, *voel* ‘feel’, and *ruik* ‘smell’ as IPP verbs, but they were not retrieved in the dataset. Robbers (1997) encountered one instance of *voel* ‘feel’ in her corpus study. The non-occurrence of *voel* and *ruik* in the *Taalkommissie* and *Wikipedia* corpora may be due to data sparseness, but for *kry* this is unlikely. *Kry* consistently occurs as a past participle in combination with a verbal complement. *Weet* ‘manage’ is the only verb encountered in the dataset that is not mentioned in the literature.

3.2 Pseudo-coordination

A second corpus investigation concerns IPP constructions with pseudo-coordination. We looked for similar constructions as in section 3.1, but with the conjunction *en* ‘and’ between the two verbs that precede or follow the auxiliary *het* ‘have’. This resulted in more than 1,900 hits, but after filtering out the false positives, only 248 examples were retained. Table 4 presents the results. Although the literature claims *en* is optional, we did not find any occurrences without *en*, neither with nor without IPP.²²

²² According to Robbers (1997) such constructions are grammatical, see example (16). In the *Taalkommissie* corpus we found a number of examples in present tense constructions, e.g. *Die ouer span staan praat in groupies* ‘The older team was talking in small groups’, but none in combination with *het*.

Table 4: IPP in constructions with pseudo-coordination in the Taalkommissie corpus and Wikipedia

Verb	Taalkommissie		Wikipedia		Total	% IPP
	IPP	PSP	IPP	PSP		
<i>sit</i> ‘sit’	48	59	2	1	110	45.45
<i>staan</i> ‘stand’	45	47	3	2	97	49.48
<i>lê</i> ‘lie’	30	5	0	0	35	85.71
<i>loop</i> ‘run’	1	5	0	0	6	16.67
Total	124	116	5	3	248	52.02

The counts in Table 4 show that IPP is optional in constructions with pseudo-coordination. This is in accordance with the literature on these constructions. The verbs occurring in constructions with pseudo-coordination are all aspectual verbs which refer to posture (*sit*, *staan* and *lê*) or movement (*loop*).²³ The verb *lê* ‘lie’ clearly prefers IPP (about 85% of the constructions), but for *sit* ‘sit’, *staan* ‘stand’ and *loop* ‘walk’ the PSP form is a valid alternative, as less than 50% of the constructions appear with IPP. Some examples are given in (48)–(49).

- (48) Sy klap die boek toe waarin sy lê en lees het, en
she* smacked the book closed in-which she lie.INF and read.INF has, and
draai na Mara toe.
turned to Mara to
‘She closed the book which she was reading and turned to Mara.’
- (49) En ons almal het gesit en wag op die droogte wat voorspel
and we all have sit.PSP and wait.INF on the drought which predicted
is...
is
‘And we were all waiting for the drought that was predicted ...’

²³ Broekhuis/Corver (2015) call this set of verbs *semi-aspectuals*, in order to differentiate them from the *core* aspectual verbs, which have different characteristics in Dutch. Considering the fact that the aspectual verbs mentioned in this section may appear in constructions with pseudo-coordination, as opposed to the aspectual verbs mentioned in section 3.1, the distinction makes sense for Afrikaans as well (cf. section 4).

3.3 Passive constructions

As mentioned in section 2, Afrikaans allows for the passivization of IPP constructions. We have encountered some examples in the data for 5 out of the 17 verbs mentioned in Table 3. Table 5 presents an overview.

Table 5: Passive IPP constructions in the Taalkommissie corpus and Wikipedia

Verb	Taalkommissie corpus			Wikipedia			Total	% IPP
	word + IPP	wees + IPP	PSP	word + IPP	wees + IPP	PSP		
<i>laat</i> 'let'	44	97	0	7	78	0	226	100.00
<i>begin</i> 'begin'	82	0	0	14	15	0	111	100.00
<i>probeer</i> 'try'	9	3	0	1	1	0	14	100.00
<i>kom</i> 'come'	0	1	0	0	0	0	1	100.00
<i>help</i> 'help'	0	1	0	0	0	1	2	50.00
Total	135	102	0	22	94	1	354	99.72

The figures in Table 5 show that passive IPP constructions are far less frequent compared to their active variants, but we find instances both in the present (with auxiliary *word*), cf. (50) and in the past (with auxiliary *wees*), cf. (51).

- (50) Indien die voorwerp *laat* val word, val dit onder die
 if the object let.INF drop.INF is, drop this under the
 invloed van gravitasie vloer toe.
 influence of gravitation floor to
 'If the object is dropped, it falls to the floor due to gravitation.'
- (51) ... 'n meganistiese hooftendens waarin alle fisiese verskynsels
 ... a mechanistic main-tendency in-which all physical phenomena
 konsekwent *probeer* herlei is tot 'n kinematiese perspektief.
 consistently try.INF reduce.INF is to a kinematic perspective
 '... a principal mechanistic tendency in which (physicists) consistently tried
 to reduce all physical phenomena to a kinematic perspective.'

In passive constructions, IPP is obligatory for the majority of the verbs. The only instance in which *help* 'help' does not have IPP is given in (52).

- (52) die Nowikof-telegram wat deur die Sowjet-ambassadeur aan die VSA
the Novikov-telegram which by the Soviet-ambassador to the USA
gestuur is, maar deur Wjatsjeslaf Molotof gelas en *gehelp*
sent is but by Vyacheslav Molotov order.PSP and help.PSP
skryf is...
write.INF is
'the Novikov telegram, which was sent by the Soviet ambassador to the
USA, but was ordered and helped to be written by Vyacheslav Molotov...'

Robbers (1997) claims that it is also possible to passivize constructions with *ophou* 'stop', but we did not find any corpus evidence for that verb. Furthermore, we did not encounter any passive constructions with pseudo-coordination in the corpus.

3.4 Constructions with modal verbs

As explained in section 2.3, Afrikaans modal verbs have a morphologically marked past tense. We have searched the *Taalkommissie* and Wikipedia corpora for past constructions with modal verbs, taking into account the alternatives with auxiliary *het* 'have' in main and subordinate clauses. We have limited the queries to constructions with adjacent verbs. Table 6 presents the filtered results for the *Taalkommissie* corpus and Wikipedia, in which the canonical $\text{pret}_1\text{-inf}_2$ constructions (type *kon werk*) are separated from the alternatives with auxiliary *het* (e.g. *het kon werk*, *kan gewerk het*). Note that '–' is used for principled non-occurrence and '0' for accidental non-occurrence.

Table 6: Constructions with modal verbs in the *Taalkommissie* corpus and Wikipedia

Verb	<i>Taalkommissie</i>		Wikipedia		Total	% canonical
	Canonical	with <i>het</i>	Canonical	with <i>het</i>		
<i>kan</i> 'can'	19,912	817	5,325	264	26,318	95.89
<i>sal</i> 'will'	16,310	1,095	6,542	220	24,167	94.56
<i>moet</i> 'must'	9,900	647	2,978	101	13,626	94.51
<i>wil</i> 'want'	7,164	254	972	21	8,411	96.73
<i>mag</i> 'may'	6	48	6	17	77	15.58

Verb	Taalkommissie		Wikipedia		Total	% canonical
	Canonical	with <i>het</i>	Canonical	with <i>het</i>		
<i>hoef</i> ‘need to’	–	14	–	2	16	0.00
<i>behoort</i> ‘ought to’	–	1	–	0	1	0.00
Total	53,292	2,876	15,823	625	72,616	95.18

The figures in Table 6 show that more than 95% of the constructions are instances of the canonical construction, which is in both the *Taalkommissie* corpus and *Wikipedia* clearly preferred over the alternatives with auxiliary *het*. An example with *wou* ‘wanted’ is given in (53).

- (53) wat ek *wou* vra, volgens watter resep kook jy jou
 what I want.PRET ask.INF according-to which recipe cook you your
 boontjiesop?
 bean-soup
 ‘What I wanted to ask, what recipe do you use for your bean soup?’

Mag ‘may’ only occurs in 15% of the cases in the canonical preterite construction. Alternatives with *het* are more common (54), but note that the verb is very uncommon in comparison to the other modals.

- (54) Baie leerders besef eers in die eksamenkamer dat, alhoewel hulle baie
 many learners realize first in the exam-room that, although they very
 hard *mag* studeer *het*, hulle nie die uitkomst bereik het
 hard may.PRES study.INF have.INF, they not the results reached have
 wat vir die module gestel is nie.
 which for the module put is not
 ‘Despite the fact they might have studied very hard, a lot of learners only
 realize in the exam room they haven’t reached the outcome that was
 expected for the module.’

The verbs *hoef* ‘need to’ and *behoort* ‘ought to’ do not have a preterite form and, as a consequence, cannot occur in the canonical construction. Their number in the corpus results is very small, as those verbs are only used in formal contexts.

In order to take a closer look at the alternative constructions, the results with *het* are split up with respect to the form and the order of the verbs in Table 7. The

ambiguous constructions ('pres-inf/pp-aux' and 'pret-inf/pp-aux') are put in separate columns.

Table 7: Constructions with modal verbs and auxiliary *het* in the Taalkommissie corpus and Wikipedia

<i>Taalkommissie</i> corpus	pret1–pp3–aux2 <i>kan gewerk het</i>	ambiguous pret–inf/pp-aux <i>kan gebeur het</i>	aux1–pret2–inf3 pret2–inf3–aux1 <i>het kon werk</i>	pres1–pp3–aux2 <i>kan gewerk het</i>	ambiguous pres–inf/pp-aux <i>kan gebeur het</i>	aux1–pres2–inf3 pres2–inf3–aux1 <i>het kan werk</i>	Total
<i>kan</i> 'can'	552	249	5	10	1	0	817
<i>sal</i> 'will'	794	275	15	8	3	0	1,095
<i>moet</i> 'must'	456	119	5	61	6	0	647
<i>wil</i> 'want'	228	14	3	4	1	4	254
<i>mag</i> 'may'	0	0	0	38	8	2	48
<i>hoef</i> 'need to'	0	0	0	14	0	0	14
<i>behoort</i> 'ought to'	0	0	0	1	0	0	1
Total	2,030	657	28	136	19	6	2,876
%	70.58	22.84	0.97	4.73	0.66	0.21	100.00

Wikipedia	pret1–pp3–aux2 <i>kan gewerk het</i>	ambiguous pret–inf/pp-aux <i>kan gebeur het</i>	aux1–pret2–inf3 pret2–inf3–aux1 <i>het kon werk</i>	pres1–pp3–aux2 <i>kan gewerk het</i>	ambiguous pres–inf/pp-aux <i>kan gebeur het</i>	aux1–pres2–inf3 pres2–inf3–aux1 <i>het kan werk</i>	Total
<i>kan</i> 'can'	184	71	8	0	0	1	264
<i>sal</i> 'will'	147	65	4	3	1	0	220
<i>moet</i> 'must'	80	15	1	4	1	0	101
<i>wil</i> 'want'	19	2	0	0	0	0	21
<i>mag</i> 'may'	0	0	0	9	7	1	17
<i>hoef</i> 'need to'	0	0	0	2	0	0	2
<i>behoort</i> 'ought to'	0	0	0	0	0	0	0
Total	430	153	13	18	9	2	625
%	68.80	24.48	2.08	2.88	1.44	0.32	100.00

Table 7 shows that all constructions mentioned in the literature were encountered, but their frequency differs markedly. The constructions can be divided into two main sets: the constructions in which the modal verb appears in its preterite (pret) form (lefthand side of the table), and the ones in which it appears in its simple present (pres) tense form (righthand side of the table). The results show that constructions belonging to the former set are more common in both corpora.

Around 70% of the alternatives are instances of pret₁-psp₃-aux₂ constructions (type *kon gewerk het*), which makes it the most common alternative for the canonical construction. An example is given in (55).

- (55) Veel langer sou sy in ieder geval nie hier *kon* *gebly*
 much longer would she in each case not here can.PRET stay.PSP
het *nie*.
 have.INF not
 ‘She would not have been able to stay here much longer anyway.’

By contrast, the constructions in which the auxiliary *het* selects a preterite modal verb (type *het kon werk*), only account for 0.97% of the constructions in the *Taalkommissie* corpus and 2.08% in Wikipedia. The majority of the constructions that are ambiguous between those constructions (i.e. the pret-inf/psp-aux cases) are, hence, instances of the former.²⁴

In comparison to the constructions with a preterite modal, the alternatives with a present form of the modal verb are uncommon. The pres₁-psp₃-aux₂ instances (type *kan gewerk het*) account for 4.73% of the constructions in the *Taalkommissie* corpus, but they only make up 2.88% of the constructions in Wikipedia. The least common construction is the one that is most common in past constructions without modal verbs, i.e. the IPP constructions (aux₁-pres₂-inf₃). They account for less than 0.3% of the constructions in both corpora. An example with *mag* ‘may’ was given in (50).

As the modals *hoef* ‘need to’ and *behoort* ‘ought to’ do not have a preterite form, we would expect them to occur as IPP verbs, but instead they consistently occur in pres₁-pp₃-aux₂ constructions. An example with *hoef* is given in (56).

²⁴ Some of the ambiguous verbs such as *probeer* ‘try’, *hanteer* ‘handle’ and *fouteer* ‘error’ are used as past participles without *ge-* in present-day Afrikaans and mentioned as such in the main prescriptive spelling dictionary (Van Huyssteen et al. 2017) and also for the first time in the latest edition of the most commonly used monolingual dictionary (Luther (ed.) 2015), although not everyone accepts it as grammatical.

- (56) Ek is trots daarop dat ek nog nooit in my lewe op iemand anders
 I am proud there-off that I yet never in my life on someone else
hoef te gereken het nie.
 need.INF to count.PSP have not
 'I am proud of the fact that never in my life have I had to count on anyone
 else.'

Also in the constructions with a present modal there are a number of ambiguous cases (pres-inf/psp-aux). They most likely belong to the pres₁-psp₃-aux₂ category.

Summing up, the corpus results show that the canonical pret₁-inf₂ construction is by far the most popular to express the past, followed by the pret₁-pp₃-aux₂ construction. Considering the alternative constructions with auxiliary *het*, it is clear that constructions in which the modal verb appears as the finite verb are preferred over constructions with an infinitival form of the modal (selected by auxiliary *het*). In combination with the fact that the use of the preterite form of the modal is preferred over the present form, IPP ends up as the least preferred option to express the past in these cases.

It would be interesting to investigate the factors that trigger the use of the different alternative past constructions, but this is left for future work.

Our findings confirm Robbers' (1997) claim that the most common construction is the canonical one, and that the instances in which the auxiliary *het* 'have' is the finite verb hardly occur.

3.5 Differences between the two corpora

If we compare the proportions of the constructions described in section 3.1 to 3.3, we see that for most verbs the proportion of IPP to no IPP is similar in both corpora. However, if we take another look at Table 3, we see that the ratio of IPP constructions is 105.2 occurrences per million words in the *Taalkommissie* corpus and 147.8 occurrences per million words in Wikipedia. A chi-square goodness of fit test reveals that there is a statistically significant difference with respect to IPP in the two corpora.²⁵

Also on the level of the individual verbs there are some noteworthy differences between the two corpora. Only the *Taalkommissie* corpus contains examples with

²⁵ The χ^2 value is 169.606. The P-value is < 0.001. The result is significant at $p \leq 0.01$.

a participle in the case of *kom* ‘come’, *bly* ‘stay’, *ophou* ‘stop’, *probeer* ‘try’, *durf* ‘dare’, *laat* ‘let’, *maak* ‘make’, and *leer* ‘teach’.²⁶

With respect to pseudo-coordination we observe that Wikipedia has significantly fewer constructions with pseudo-coordination than the *Taalkommissie* corpus.²⁷

The differences between the corpora are most likely due to the inclusion of nonstandard Afrikaans in the *Taalkommissie* corpus, but as we do not have any metadata directly linked to the sentences in the corpus, it is hard to investigate this in a systematic way.

3.6 Conclusion

The corpus study has revealed some interesting insights with respect to the occurrence of IPP in Afrikaans. With respect to the double infinitive constructions, we can conclude that verbs that may appear in such constructions are obligatory IPP verbs for most speakers. Within this set, the aspectual verbs turn out to be the most common IPP verbs. A subset of the verbs that appear in double infinitive construction may also occur as IPPs in passive constructions.

With respect to constructions with pseudo-coordination, the corpus results confirm the statement in the literature that IPP is optional.

Modal verbs have a special status. For those verbs IPP seems the least preferred option in comparison to alternative ways to express the past. Instead, a construction in which they appear as a preterite finite verb selecting a bare infinitive is canonical.

4 Integrating Afrikaans in the typology of IPP verbs

As already observed in the introduction, Afrikaans shares the IPP phenomenon with German and Dutch. For a study of the similarities and differences with these

²⁶ For each verb in Table 3, a Fisher’s exact test (Fisher 1922) was performed, as it is able to deal with small sample sizes, but only for *leer* ‘learn’ was the difference statistically significant (the Fisher exact test statistic value is 0.0311; the result is significant at $p < 0.05$).

²⁷ Again, a goodness of fit test was performed. The χ^2 value is 19.846. The P-value is < 0.001 . The result is significant at $p \leq 0.01$.

languages we employ the typology of IPP verbs in Augustinus/Van Eynde (2017). This typology is based on a usage-based analysis of Dutch and German treebanks, i.e. Lassy Small and CGN core for Dutch and Tüba-D/S and Tüba-D/Z for German. For Afrikaans the corpus search yielded 8,122 examples with IPP, see Table 8 for an overview. Comparing this to the data from Dutch and German IPP, it is immediately clear that the phenomenon is less frequent in Afrikaans, see Table 9.

Table 8: IPPs in the Taalkommissie corpus and Wikipedia

Construction	<i>Taalkommissie</i> corpus	Wikipedia	Total
Double infinitive	5,681	1,951	7,632
Pseudo-coordination	124	5	129
Passive	237	116	353
Modal	6	2	8
Total	6,048	2,074	8,122

Table 9: IPP per 10,000 words

	IPPs/words in corpus	IPPs per 10K words
Dutch	1,101 / 1,983,788	5.55
German	400 / 1,468,415	2.72
Afrikaans	8,122 / 67,088,350	1.21

To some extent this is due to the fact that the numbers for Afrikaans are a slight underestimation given that we only consider adjacent verbs and verbs which are separated by at most one other word. The main factor, though, is the very low frequency of IPP forms for the modals. While they are the most frequently used IPP verbs in the Dutch and German treebanks, the Afrikaans corpora contain just a few examples, see (54). If the modals are left out of the comparison, Afrikaans takes an intermediate position, having fewer IPP forms than Dutch, but more than German.

A finer-grained comparison can be made on the basis of the typology in Figure 1.

Fig. 1: A typology of IPP in Dutch, German and Afrikaans

The percentages for Dutch and German are quoted from Augustinus/Van Eynde (2017). They concern the relative frequency in the respective treebanks. The percentages for Afrikaans are based on the corpus study in section 3.

Within the class of the modal verbs we make a distinction between core modals and evidential modals. The former are, as discussed previously, the most frequently used IPPs in Dutch and German, but they are nearly absent in Afrikaans. The few exceptions include some hits for *mag* ‘may’, *wil* ‘want’ and *kan* ‘can’. The evidential modals are rarely combined with a perfect auxiliary in any of the three languages. If they do, they occasionally take the IPP form in Afrikaans, as in (40). This is also possible in Dutch, but in order to attest it Augustinus/Van Eynde (2017) had to resort to internet search, since the treebanks did not contain any hits.

Within the class of the aspectual verbs we make a distinction between the core aspectuals and the semi-aspectuals.²⁸ The distinction is relevant for Afrikaans, since the former turn up in the double infinitive construction while the latter turn up in pseudo-coordination. It is also relevant for Dutch, since the core aspectuals take *zijn* ‘be’ as the auxiliary of the perfect, while the semi-aspectuals take *hebben* ‘have’, cf. *Ik ben gaan lopen* ‘I went and ran’ versus *Ik heb staan praten* ‘I was talking’. The odd one out in this category is German. While the aspectuals account for more than half of the IPP verbs in Afrikaans and for nearly 30% of the IPP verbs in Dutch, the German aspectual verbs do not take the IPP form.

The subject control IPP verbs are a heterogeneous group. There is one that occurs in all three of the languages, i.e. *leer/leren/lernen* ‘learn’, and there are three that occur in Afrikaans and Dutch, but not in German, i.e. *probeer/proberen* ‘try’, *durf/durven* ‘dare’ and *weet/weten* in the sense of ‘manage to’.

²⁸ The term ‘semi-aspectual’ is adopted from Broekhuis/Corver (2015: 151).

The causative IPP verbs are few in number but relatively high in frequency: *laat* and *maak* jointly account for 27.54% of the Afrikaans IPP verbs, *lassen* for 22.5% of the German IPP verbs, and *laten* and *doen* for 14.53% of the Dutch IPP verbs.

The perception IPP verbs are the same in the three languages, i.e. *sien/zien/sehen* ‘see’ and *hoor/horen/hören* ‘hear’. Examples with *voel/voelen* ‘feel’ can be constructed and/or googled, but are not attested in the corpora that we consulted.

The benefactive IPP verbs are *help/helpen/hilfen* ‘help’ and *leer/leren* in the sense of ‘teach’. They account for a very small percentage of the IPP verbs. For German we did not find a single hit in the treebanks.

In contrast to Dutch and German, Afrikaans also allows the use of IPPs in combination with the auxiliaries of the passive, i.e. *word* and *wees*. The instances that we found in the corpora mainly concern the causative *laat* (226 hits), the aspectual *begin* (111 hits) and the control verb *probeer* (14 hits).

The resulting classification can be compared to the one in Schmid (2005), which also covers Dutch, German and Afrikaans, as well as some regional variants of Dutch and German, i.e. Bernese German, Sankt Gallen German, Zürich German and West Flemish. Schmid distinguishes eight classes of IPP verbs, just as we do, but the details of the classification are different for the subject-oriented ones. Within the class of aspectuals, she differentiates between durative and inchoative, which in our classification are both core aspectuals, and she does not include the semi-aspectuals, which is justified for German, but not for Dutch or Afrikaans. The evidential modals are called raising verbs in Schmid (2005), which is a misnomer, since the modals and most of the aspectuals are raising verbs as well.

As shown in Augustinus/Van Eynde (2017), Schmid’s claims about the status of IPP for the eight classes (obligatory, optional or impossible) is in line with the judgments in the literature and with corpus data for German, except in the case of the causative *lassen* ‘let’, which she classifies as an obligatory IPP verb, while it is in fact an optional IPP verb, as claimed in Duden (2006) and confirmed by the corpus data discussed in Augustinus/Van Eynde (2017). For Afrikaans, Schmid claims that IPP is optional for all classes, except for the raising verbs, which according to her do not allow it. This is contradicted by the fact that *blyk* ‘seem’ does occur as IPP in the corpora we consulted. The characterization of the other classes as optional IPP verbs is not by itself erroneous, but it is misleading since it does not differentiate between classes where the IPP form is by far the most frequent (the double infinitive construction), classes where it is more or less in 50–50 relation with the non-IPP form (pseudo-coordination) and classes where it is marginal (the modals).

5 Conclusion

We conducted a corpus study of two Afrikaans corpora in order to identify the verbs showing the IPP effect. The corpus results reveal that the majority of the IPP constructions concern instances of the double infinitive construction. In addition, Afrikaans IPPs also occur in constructions with pseudo-coordination and in passive constructions.

Next, we compared the results to a previous corpus study on IPP in German and Dutch. IPP is more widespread in Afrikaans than in Dutch and German, as in those languages IPP is only possible in the double infinitive. In order to compare the verb types that figure in IPP constructions, we extended the classification in Augustinus/Van Eynde (2017) on Dutch and German IPP to Afrikaans. With respect to the verb types that may appear as IPP, Afrikaans resembles Dutch, as the IPP shows up in the same verbal categories (i.e. modal, aspectual, subject control, causative, perception and benefactive). The relative frequencies of the corpus study indicate, however, that the two languages highly differ with respect to the category of core modal verbs. While this is the prototypical IPP category in Dutch, Afrikaans modals hardly show the IPP effect. Instead they canonically use an alternative construction to express the past. The prototypical IPP category in Afrikaans is the set of aspectual verbs. If we compare Afrikaans to German, we see that German aspectual verbs never occur as IPPs.

In future work we aim to create a large treebank for Afrikaans. Due to the limitations of the corpora, we had to limit our queries to constructions in which the verbs were (almost) adjacent. Using a (large scale) treebank would allow the investigation of constructions with non-adjacent words in a more systematic way.

References

- Augustinus, Liesbeth/Dirix, Peter (2013): The IPP effect in Afrikaans: A corpus analysis. In: Oepen, Stephan/Hagen, Kristin/Bondi Johannessen, Janne (eds.): *Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa 2013)*, May 22–24, 2013, Oslo, Norway. (= *Linköping Electronic Conference Proceedings* 85). Linköping: Linköping University Electronic Press. 213–225.
- Augustinus, Liesbeth/Dirix, Peter/Van Niekerk, Daniel/Schuurman, Ineke/Vandeghinste, Vincent/Van Eynde, Frank/Van Huyssteen, Gerhard (2016): *AfriBooms: An online treebank for Afrikaans*. In: *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)*, Portorož. Paris: European Language Resources Association (ELRA). 677–682.
- Augustinus, Liesbeth/Van Eynde, Frank (2017): A usage-based typology of Dutch and German IPP verbs. In: *Leuvense Bijdragen: Tijdschrift voor Germaanse Filologie* 101. 101–122.

- Breed, Catharina Adriana (2012): Die grammatikalisering van aspek in Afrikaans: Semantiese studie van perifraستiese progressiewe konstruksies. PhD thesis. Potchefstroom: North-West University.
- Broekhuis, Hans/Corver, Norbert (2015): *Syntax of Dutch: Verbs and verb phrases*, Vol. II. Amsterdam: Amsterdam University Press.
- Conradie, Charl Jacobus (2006): Werkwoordsclusters in contrast: Het Nederlands en het Afrikaans. In: *Colloquium Neerlandicum* 16. 247–260.
- CText (2011): *Taalkommissie corpus*, version 1.0.
- Den Besten, Hans/Edmondson, Jerold A. (1981). The verbal complex in continental West-Germanic. In: Abraham, Werner (ed.): *On the formal syntax of the Westgermania*. Papers from the 3rd Groningen Grammar Talks, Groningen, January 1981. (= *Linguistik aktuell* 3). Amsterdam/Philadelphia: Benjamins. 155–216.
- De Vos, Mark (2001): *Afrikaans verb clusters: A functional-head analysis*. Master's thesis. Tromsø: University of Tromsø.
- De Vos, Mark (2005): *The syntax of pseudo-coordination in English and Afrikaans*. Utrecht: Landelijke Onderzoekschool Taalwetenschap.
- Dirix, Peter/Augustinus, Liesbeth/Van Niekerk, Daniel/Van Eynde, Frank (2017): Universal Dependencies in Afrikaans. In: de Marneffe, Marie-Catherine/Nivre, Joakim/Schuster, Sebastian (eds.): *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, 22 May 2017, Gothenburg, Sweden. (= *Linköping Electronic Conference Proceedings* 135). Linköping: Linköping University Electronic Press. 38–47.
- Donaldson, Bruce C. (1993): *A grammar of Afrikaans*. (= *Mouton Grammar Library* 8). Berlin: Mouton de Gruyter.
- Duden (2006): *Der Duden in zwölf Bänden*. Vol. 4: *Die Grammatik*. Unentbehrlich für richtiges Deutsch. 7th edition. Mannheim et al.: Dudenverlag.
- Fisher, Ronald A. (1922): On the interpretation of χ^2 from contingency tables, and the calculation of P. In: *Journal of the Royal Statistical Society* 85, 1. 87–94.
- Luther, Jana (ed.) (2015): *Handwoordeboek van die Afrikaanse Taal (HAT)*. 6th edition. Cape Town: Pearson.
- Müller, Stefan (2002): *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German*. (= *Studies in Constraint-based Lexicalism* 13). Stanford: CSLI.
- Pilon, Suléne (2005): *Outomatiese Afrikaanse woordsoortetikettering*. Master's thesis. Potchefstroom: North-West University.
- Ponelis, Friedrich Albert (1979): *Afrikaanse sintaksis*. Pretoria: Van Schaik.
- Ponelis, Friedrich Albert (1993): *The development of Afrikaans*. Frankfurt a.M.: Lang.
- Robbers, Karin (1997): *Non-finite verbal complements in Afrikaans: A comparative approach*. (= *HIL Dissertations* 33). The Hague: Holland Academic Graphics.
- Schlünz, Georg Isaac (2010): *The effects of part-of-speech tagging on text-to-speech synthesis for resource-scarce languages*. Master's thesis. Potchefstroom: North-West University.
- Schmid, Tanja (2005): *Infinitival syntax. Infinitivus pro participio as a repair strategy*. (= *Linguistics Today* 79). Amsterdam/Philadelphia: Benjamins.
- Stell, Gerald (2011): *Ethnicity and language variation: Grammar and code-switching in the Afrikaans speech community*. (= *Research in African Studies* 19). Frankfurt a.M.: Lang.
- Van Huyssteen, Gerhard/Hendricks, Frank/Pilon, Suléne/Beyer, Herman/Kapp, Sophia/Lombard, Frikkie/Louw, Phillip/Luther, Jana/Pienaar, Marné (2017): *Afrikaanse Woordelys en Spelreëls (AWS)*. 11th edition. Cape Town: Pharos.
- Van Schoor, J.L. (1983): *Die grammatika van Standaard-Afrikaans*. Cape Town: Lex Patria.

- Verdoolaege, Annelies/Van Keymeulen, Jacques (2010): *Grammatica van het Afrikaans*. Ghent: Academia Press.
- Wurmbrand, Susi (2004): West Germanic verb clusters: The empirical domain. In: Kiss, Katalin É./van Riemsdijk, Henk (eds.): *Verb clusters: A study of Hungarian, German, and Dutch*. (= *Linguistics Today* 69). Amsterdam/Philadelphia: Benjamins. 43–85.
- Zwart, Jan-Wouter (2007): Some notes on the origin and distribution of the IPP-effect. In: *Groninger Arbeiten zur germanistischen Linguistik* 45. 77–99.