Abstract

Khoekhoe, a Central Khoisan language, has been claimed to have a clause-second position and topological fields similar to German and Dutch. The position in front of the clause-second position can be occupied by either the matrix verb or a dependent. We argue that monomoraic words are exempt from the general head-final order of Khoekhoe and suggest that this can give rise to discontinuous constituents, where second-position clitics intervene within the VP. We show that this idea provides a simple account of Khoekhoe word order variation and formalize it within a linearization-based HPSG analysis that has a wider scope than the previous Minimalist analyses of Khoekhoe and that is compatible with evidence from tonology.

This paper examines word order variation in Khoekhoe (short for Khoekhoegowab, also known as Nama/Damara), a Central Khoisan language spoken in Namibia and South Africa. It has been claimed to have a clause-second position (den Besten, 2002) and to have topological fields similar to German and Dutch (Witzlack-Makarevich, 2006). There have been three previous generative analyses in GB/Minimalism (Washburn, 2001, den Besten, 2002, Huybregts, 2003). As these analyses seem to be based entirely on descriptive grammars, they are based on a limited amount of data and, while capturing essential aspects of Khoekhoe word order, make some empirical assumptions that turn out to be problematic in the light of corpus data and elicited data. Our goal will be to use new elicited data, supported by corpus data, to give a more complete picture of the empirical situation and then develop a comprehensive analysis, which we will formalize in HPSG.

1 Basic Data

1.1 SOV Clauses

Khoekhoe is an SOV language (1a).[1] In matrix clauses, one of the clause type markers *ge* (declarative), *kha* (interrogative), *ko/km* (assertive) may appear immediately after the subject. Embedded verbs, stripped of all their non-clitical arguments, the negation marker *tama*, and pronominal object clitics attach to the verb, forming what we will call the *verbal complex*. The linearization of the elements between the clause type marker and the verbal complex is largely free. In (1a), all six permutations of these elements are grammatical (e.g., 1b–d). By analogy

[†]I want to thank Berthold Crysmann, Stefan Müller, two anonymous reviewers, and the audience of HPSG 2013 for their helpful comments. Most of all, I thank Gerson Topnaar for sharing his language with me. Of course, I alone am responsible for all errors.

[1]The transliteration largely follows the official transliteration as given in Haacke & Eiseb (2002), with two differences in vowel marking: First, long vowels, including nasalized ones, are marked by reduplication. Second, epenthetic vowels are omitted, following the analyses of Hagman (1977) and Brugman (2009). As in the official transliteration, tone is omitted.

to Germanic, we will refer to the collection of these elements as the *Middlefield*, following Witzlack-Makarevich (2006):

(1)

			Middlefield			
a.	taras	(ge)	ǁari	ǂkhanis-a	go	maa=te
	woman	DECL	yesterday	book-A	TAM	give=me
b.	taras	(ge)	ǂkhanisa	ǁari	go	maa=te
c.	taras	(ge)	ǁari	go	ǂkhanisa	maa=te
d.	taras	(ge)	go	ǁari	ǂkhanisa	maa=te

'the woman gave me a book yesterday'

As these examples show, complement NPs are realized as full noun phrases with the case marker *-a* (glossed A) or as a pronominal clitic in the verbal complex. A subject may be realized by a full NP as in (1), or by an enclitic Person-Gender-Number (PGN) marker appearing in the normal subject position (2a). In the latter case, a coreferent case-marked NP may appear in the middlefield (2b):

(2) a. tsîî=s ge ǁari ǂkhanis-a go maa=te
and=3fs DECL yesterday book-A TAM give=me
'and she gave me a book yesterday'

b. tsîî=s ge ǁari taras-a ǂkhanis-a go maa=te
and=3fs DECL yesterday woman-A book-A TAM give=me
'and the woman gave me a book yesterday'

With respect to case marking and word order, this NP behaves largely like a complement (cf. Haacke (1978) for an extensive discussion). We will assume that it is indeed a complement introduced by a lexical rule and will not consider it further here.

The subject PGN marker is a clitic rather than an affix, as it does not seem to show any morphophonological idiosyncrasies (Zwicky & Pullum, 1983), and may attach to any element preceding the subject position, even if it does not belong to the same clause. In particular, it may appear at the beginning of embedded clauses:

(3) tsîî=[b |gôab-a !narigau] hîa=gu ge ǁîga go !gûû[2]
and=3MS boy-A drive while=3mp DECL they.A TAM go
'and while the boy was driving, they left'

As it apparently can have wide scope over VP coordination only if the subject is the same in both VPs, we analyze the PGN marker as a subject pronoun, not an agreement marker.

[2] Witzlack-Makarevich (2006, 57)

1.2 Fronting

If the subject is realized by a clitic only, a complement (4a), an adjunct (4b), or the verbal complex itself (4c) may precede it. In declaratives, there may be at most one such element (5). The subject position cannot be occupied by a lexical NP in this case (6):

(4)

	Prefield			*Middlefield*		
a.	ǂkhanis-a	=s	ge	go	‖ari	maa=te
	book-A	3fs	DECL	TAM	yesterday	give=me
b.	‖ari	=s	ge	go	ǂkhanis-a	maa=te
c.	maa=te	=s	ge	go	‖ari ǂkhanis-a	

'she gave me a book yesterday'

(5) * ǂkhanis-a maa=te=s ge go ‖ari
book-A give=me=3FS DECL TAM yesterday
intended: 'she gave me a book yesterday'

(6) * ǂkhanis-a taras ge go ‖ari maa=te
book-A woman DECL TAM yesterday give=me
intended: 'the woman gave me a book yesterday'

Again by analogy with Germanic V2, we refer to the preposed element as the *Prefield* and refer to the occurence of an element in the prefield as *fronting*. The position of the subject clitic and the clause type marker has been interpreted as a clause-second position by den Besten (2002) and Witzlack-Makarevich (2006).

The verbal complex can be fronted as a whole, while it is impossible to front only parts of the verbal complex. While the prefield may contain at most one element in declaratives, it may contain several elements in interrogative and hortative clauses (7). Any collection of elements that could occur alone in the prefield appears to be allowed. If the verbal complex is fronted, it is the last element of the prefield (7 d–e):

(7)

		Prefield				*Middlefield*		
a.		tarasa	ǂkhanisa	maa	=b	go	‖ari?	
		woman	book	give	=3MS	TAM	yesterday	
b.		tarasa	ǂkhanisa		=b	go	‖ari	maa?
c.		ǂkhanisa	tarasa		=b	go	‖ari	maa?
d.	*	tarasa	maa	ǂkhanisa	=b	go	‖ari?	
e.	*	maa	tarasa	ǂkhanisa	=b	go	‖ari?	

'did he give the woman a book yesterday?'

The clause type marker *kha* may appear between any two prefield elements (e.g., 8), while it may never appear in the middlefield.

(8) taras-a kha maa=ts go ǂkhanis-a ǁari?
woman-A INTERROG give=2MS TAM book-A yesterday
'did you give the woman a book yesterday?'

Previous work mentions multiple fronting only in the case where there is a clitic-left-dislocated subject and the last fronted element is the verbal complex (Hagman, 1977, 144). Indeed, this is by far the most common pattern, but there is corpus data supporting the more inclusive judgments in (7). Multiple fronting without a dislocated subject is attested at least if the subject is in the first or second person.[3] Possibly, multiple fronting requires the subject to be topical. In any case, it seems that the requirement of a dislocated subject is a matter of discourse constraints, not of grammaticality. Multiple fronting where the last element is not the verbal complex is indeed very rare and we have only been able to find a handful of corpus examples.[4]

2 More Data

TAM Markers Kheokhoe TAM markers fall into two classes with differing word order possibilities: monomoraic TAM markers (*a, ga, ge, go, nî, ra*), and multimoraic markers (*tide, ii, hââ*). In the examples above, all TAM markers were monomoraic. Monomoraic TAM markers often immediately precede the verbal complex, but they can also be placed further to the left within the middlefield (1c–d).[5] They may also occur immediately after the verbal complex.[6] Thus, monomoraic TAM markers may occur anywhere from the beginning of the middlefield to the position following the verbal complex, subject to certain usage preferences. On the other hand, multimoraic markers are always placed after the verbal complex:

(9) a. namas ge taras-a maa tide
Nama(f.) DECL woman-A give TAM+NEG

[3] E.g., [xuu'e] kha [xare] [!gûûs ââts !nââ] [daa-khâi]=ts go – thing-A INTERROG ADV journey your on hurry=2ms TAM 'did you hurry on your journey?' (Krönlein, 1889, 47)

[4] E.g., [!gararo-e] [xawe'e]=ts kha uu-hâ tama hâ sa |goan xa – remainder-A ADV=2MS INTERROG have not TAM your cattle of 'do you not even have a small remainder of your cattle?' (Krönlein, 1889, 93)

[5] This is not mentioned by the descriptive grammars. Examples are given by Haacke (1999), Witzlack-Makarevich (2006), and Brugman (2009, 244). While this is mainly found in embedded clauses, it is attested in main clauses at least in older data and in Witzlack-Makarevich (2006)'s data from the Richtersveld dialect.

[6] This is again not mentioned by the descriptive grammars. Examples are given by Klein (1976, 215) and Haacke (1999, 191, S14-S17a). There is also a small amount of corpus data, e.g. *hoohoo=te=ts ge o=ta kom hoo!ûâ go=o* – warn=me=2MS TAM CONJ=1S ASSERT find.out TAM=ASSERT 'after you had warned me, I found out (about it)' (Krönlein, 1889, 109). Our consultant apparently only accepts this pattern if the TAM marker is not sentence-final.

b. * namas ge taras-a tide maa
Nama(f.) DECL woman-A TAM+NEG give
'the Nama will not give (anything) to the woman'

There are also multiple, potentially discontinuous TAM markers, such as *go...a* in the following example:[7]

(10) tsuuse=ta ra tsââ [tiita go ǁari a mîîba=tsi] !khais !aroma[8]
painful=1S TAM feel I TAM yesterday TAM tell=you thing because
'I feel bad because of the thing I told you about yesterday'

Other Dislocation Patterns There is an apparent second fronting strategy, where the verb is fronted together with the TAM markers and optionally some complements and adjuncts, while others may be realized in the middlefield (Hagman, 1977, 111). The subject position may be occupied by a full NP in this case. Its prosodic behavior (Haacke, 1999, Brugman, 2009) and a close parallelism to copulative clauses suggest that this structure is treated best as a fossilized cleft construction where the fronted elements form a constituent modifying the subject, together with which they occupy the prefield. We will not treat this structure here.

Further to the left of the prefield, there may appear dislocated NPs that are coreferent with a pronoun in the subsequent part of the clause. This fronting strategy seems to be an instance of clitic left dislocation and will not concern us further here. Apart from the local fronting of wh-prases into the prefield, there is no wh-extraction in Khoekhoe.

Tonology Khoekhoe is a tone language (Haacke, 1999). Tone is determined lexically, but the lexical tone melodies are replaced in a predictable way in certain syntactic environments by sandhi melodies. While a formal analysis of tone is beyond the scope of this paper, tone should be considered in any syntactic analysis, as the plausibility of syntactic analyses can be measured by the simplicity of the rules necessary for describing the interaction between syntax and tonology.

For every basic tonal melody, there is an associated sandhi melody (Haacke, 1976, Haacke, 1999). Informally, sandhi patterns generally occur on a word when the word is a noninitial daughter. The pattern is most transparent in NPs and PPs, where, in the simplest case, every noninitial dependent and the phrase-final head

[7] While combinations of a monomoraic marker with multimoraic markers such as *ge...ii* and combinations of the form *gV-rV* are well known, other combinations of monomoraic markers have been documented by Witzlack-Makarevich (2006, 21) for Richtersveld Nama and by Haacke (2013, 346) for !Gora, who links the *a* occuring here with the juncture morpheme of Kalahari Khoe. (10) is an example from written Namibian Khoekhoegowab.

[8] |Uriseb (1993, 4)

show sandhi tone (Haacke, 1976). As the tones inside NPs and PPs are determined entirely by the NP/PP-internal structure, and adverbs and particles do not change their tone (Brugman, 2009, 169), the syntactic constructions that concern us here only affect the tone of TAM markers and the matrix verb. Their tonal behavior, as described by Haacke (1999) and Brugman (2009), has two important ramifications for our concerns. First, it may be suggested that the fronted elements in (7a–c) form a partial VP, which would be fronted as a constituent – similar to the analysis of German multiple fronting by Müller (2002b, 2013). However, as the verb would be the last constituent of the fronted partial VP, it would then be expected to carry sandhi tone in multiple fronting, which it does not (Haacke, 1999). Thus, tonology suggests that the fronted elements in multiple fronting do not form a constituent. Second, it suggests that the verbal complex forms a constituent with TAM markers: while the verb shows sandhi tone if it follows the TAM marker, it usually shows citation tone if it is followed by a TAM marker (Haacke, 1999, 189, Brugman, 2009, 261).[9] These considerations will play a role in the formal details of our analysis.

3 Analysis

3.1 Previous work

Word order variation in Khoekhoe has previously been studied in GB and Minimalism by Washburn (2001), den Besten (2002), and Huybregts (2003). They all agree that fronting arises from movement out of the canonical position. The highest (overtly filled) head is an element that resists preposing and under the analyses is a complementizer: the clause type marker in the analyses of Washburn (2001) and den Besten (2002), and the subject clitic according to Huybregts (2003).

The three analyses seem to rely entirely on descriptive grammars, which has the unfortunate consequence that they sometimes make incorrect empirical assumptions. For instance, the presentation in Hagman's descriptive grammar (1977) suggests that TAM markers are always adjacent to the verbal complex unless it is fronted. In Washburn (2001)'s analysis, where TAM markers occupy the T position, this empirical assumption motivates the claim that complements are not overtly realized in the maximal projection of the verb, which then motivates the fact that declaratives do not allow multiple fronting (5). However, the empirical claim, while correctly expressing a general tendency, is falsified by (1c–d). Thus, there is no motivation for the ungrammaticality of (5).

Huybregts is mainly concerned with explaining a presumed pattern where the TAM marker is shifted to the beginning of the middlefield when the verbal complex is fronted. Again, this pattern actually reflects a usage preference, not a grammati-

[9] Verbs only show sandhi tone if there is a clause type marker (Brugman, 2009, 256). We assume that this rule has to be stipulated and does not follow from constituent structure.

cality contrast. Besides the empirical problems, the two analyses have gaps that are not trivial to fill. For instance, Washburn only accounts for declaratives, leaving open how (7) can be accounted for. Den Besten (2002) cosiders more (though only positive) data, which results in a far more comprehensive empirical picture. However, the analysis is rather informal and it is left open which mechanisms precisely generate the grammatical structures and how clauses like (5) are excluded (den Besten, 2002, 38).

3.2 Fronting as Extraction

The basic idea of the previous analyses, i.e. that fronting is an instance of dislocation and that clauses without a filled prefield are 'more basic', seems very reasonable. From a more theory-neutral point of view, one may take the Subject – Middlefield – Verbal Complex sequence as the basic word order, and allow elements that would appear behind the subject in the basic order to appear in front of the subject clitic, with the syntactic structure and the linearization of the remainder remaining invariant.

The difficulty arises that fronting applies equally to the verb and its dependents. In the GB/Minimalist accounts, this is no problem, as there are intermediate heads: T and C in den Besten's analysis, and many more in the Minimalist analyses. As these positions are filled by overt elements that indeed resist preposing, namely the TAM marker, the clause type marker, or the subject, the analysis appears to be well-motivated. However, as the following set of examples shows, none of the candidate heads is necessary for fronting to be allowed: a nonreferential subject clitic, the clause type marker, and the TAM marker can be omitted:

(11)

	Prefield							
a	[ǂhanu	tama]	='i	kom	hâã	[‖gâus	ai]	=o
	be.right	NEG	=3NS	ASSERT	TAM	home	in	=ASSERT
b.	[ǂhanu	tama]	='i	kom		[‖gâus	ai]	=o
c.	[ǂhanu	tama]		kom		[‖gâus	ai]	=o
d.	[ǂhanu	tama]	='i			[‖gâus	ai]	
e.	[ǂhanu	tama]				[‖gâus	ai]	

'Something is not good at home (a-c, e) / Is something not good at home? (d)'

Under an analysis along the lines suggested above, a uniform analysis of fronting can only be achieved by using empty elements, or by some other way of mimicking the effect of an empty head. Depending on one's convictions, one may either take this as evidence for empty elements, or as evidence that there might be better analyses. As we are not aware of any other phenomenon of Khoekhoe which would provide an independent motivation for empty clause type markers, TAM markers, or subject clitics, we take this as a motivation to look for alternatives.

3.3 Linearization-based Analysis

Descriptively, fronting seems to be purely a matter of constituent order and information structure. There are no known concomitant morphosyntactic effects, we are not aware of differences in semantic interpretation, and there is no particular reason to link fronting to differences in constituent structures. This suggests that one might look for a linearization-based analysis which models fronting as variation in linearization without variation in constituent structure.

Brugman (2009) observes that some of the constraints on the linearization of subject clitics and clause type markers follow from the fact that they consist of at most one mora. In her analysis, mono-moraic words are not prosodic words in Khoekhoe, and hence cannot appear in initial positions. In fact, monomoraic words more generally tend to show unexpected word order patterns in Khoekhoe, when compared to their multimoraic counterparts, which usually show head-final order: PPs and CPs are head-final, the verb usually follows complements and adjuncts, and NPs are generally head-final. The contrast between mono- and multimoraic elements is particularly clear in the case of complements: prosodically autonomous non-extraposed complements generally precede the verb, while object clitics follow the verb (1).

This is somewhat parallel to the contrast between monomoraic and multimoraic TAM markers discussed above. We suggest that TAM markers are always heads selecting a verbal complex. Monomoraic words, not being prosodic words, are exempt from the general head-final word order. Thus, both the free placement of the monomoraic markers and the obligatory postverbal position of bimoraic TAM markers, which would require additional stipulations if TAM markers were treated as complements or via a construction, follow automatically. The same treatment may be applied to the unexpected word order properties of the subject clitics and clause type markers, which are also monomoraic: they attach to a satured projection of a verb or a TAM marker, but may be realized within the VP, which thus becomes discontinuous and encompasses both the prefield and the middlefield. This immediately explains why both the predicate and its dependents can appear in the prefield, as the prefield is simply the first part of the discontinuous VP. It also explains why fronting is impossible if the subject position is filled by a lexical NP, as lexical NPs are multimoraic and hence obey strict head-final word order and cannot intervene in another phrase. This treatment is compatible with the fact that the placement of subjects and clause type markers elements is sensitive to the *number* of elements in the prefield, but insensitive to their syntactic categories.

None of these facts appears to be motivated so easily by the extraction-based analyses. Thus, it seems that a more adequate analysis of Khoekhoe word order can be provided on the basis of the claim that Khoekhoe is a head-final language, but that fronted constituents and monomoraic words are exempt from this constraint and that their deviant linearization behavior may give rise to discontinuous constituents. In the next section, we will formalize such an analysis.

4 Formalization

The intuition of discontinuous constituents can be formalized elegantly in Domain-based HPSG (Reape, 1994). Every sign is assigned a *domain* encoded in the DOM list, whose elements are of type *dom-obj*. The phonology of a phrase is computed not from the phonologies of the daughters, but from the phonologies of the domain elements:

(12)
$$\textit{sign} \rightarrow \begin{bmatrix} \text{PHON} & \boxed{1} \oplus \ldots \oplus \boxed{n} \\ \text{DOM} & \left\langle \left[\text{PHON} \quad \boxed{1}\right], \ldots, \left[\text{PHON} \quad \boxed{n}\right] \right\rangle \end{bmatrix}$$

Phrases differ as to how their domain is computed from the domains of their daughters (Kathol & Pollard, 1995, Donohue & Sag, 1999): *Compacting* phrases fuse the daughter's domains into a single domain element, while *Liberating* phrases take all domain elements of the daughters into their domain. Using liberating phrases, domain structure can be dissociated from constituent structure.

(13) a. *liberating-phrase* $\rightarrow$
$$\begin{bmatrix} \text{DOM} & \boxed{1} \circ \ldots \circ \boxed{n} \\ \text{DTRS} & \left\langle \left[\text{DOM} \quad \boxed{1}\right], \ldots, \left[\text{DOM} \quad \boxed{n}\right] \right\rangle \end{bmatrix}$$

b. *compacting-phrase* $\rightarrow$
$$\begin{bmatrix} \text{DOM} & \left\langle \left[\text{DOM} \quad \boxed{1} \circ \ldots \circ \boxed{n}\right] \right\rangle \\ \text{DTRS} & \left\langle \left[\text{DOM} \quad \boxed{1}\right], \ldots, \left[\text{DOM} \quad \boxed{n}\right] \right\rangle \end{bmatrix}$$

In Khoekhoe, the verbal complex, noun phrases, and postpositional phrases are compacting, while higher projections of the verbal complex and of TAM markers are liberating. For the purposes of our analysis, domains that only allow permutation within maximal projections, as used by Müller (1999), would also provide a viable option. Word order is described by the interaction of constituent structure with linearization constraints. We will first discuss the constituent structure of Khoekhoe clauses, and then define appropriate linearization constraints.

4.1 Constituent Structure

Verb Phrase For the VP, several constituent structures have been proposed for languages that show a similar degree of free word order. It may be flat (Kasper, 1994), binary branching, or have an intermediate structure where arguments are realized in a flat structure, while adjuncts adjoin at higher levels. We choose a binary branching structure, as it avoids technical complications of the other options and readily accounts for simple coordination phenomena. The verbal complex and the VP are built up recursively by *head-comp-phrase* and the *head-adj-phrase*:

(14)
$$\begin{bmatrix} \textit{head-comp-phrase} & \\ \textsc{synsem|loc|cat|comps} & \boxed{1} \oplus \boxed{3} \\ \textsc{head-dtr|s|l|cat|comps} & \boxed{1} \oplus \left\langle \boxed{2} \right\rangle \oplus \boxed{3} \\ \textsc{non-head-dtrs} & \left\langle \left[\textsc{synsem} \; \boxed{2} \right] \right\rangle \end{bmatrix} \quad \begin{bmatrix} \textit{head-adj-phrase} & \\ \textsc{head-dtr|synsem} & \boxed{1} \\ \textsc{non-head-dtrs} & \left\langle \left[\ldots \textsc{|mod} \; \boxed{1} \right] \right\rangle \end{bmatrix}$$

Note that the verb can combine with its complements in any order, which accounts for the free word order within the middlefield Müller (2013, To Appear).

Verbal complex Verbal complex formation is enforced by a boolean-valued feature LEX (Müller, 2002a, 87). The analysis of the Khoekhoe verbal complex is simple: Unlike, for instance, German, Khoekhoe only allows clitics and verbs to appear in the verb complex, and none of these elements can be realized outside of the verbal complex. Therefore, we can simply assume that every sub-constituent of the verbal complex is LEX +, while all higher constituents are LEX −. Clitics occuring in the verbal complex and PGN markers are lexically specified as LEX +. Raising verbs select a predicative complement marked as LEX +. NPs, adverbs, and other constituents that cannot occur in the verbal complex are specified as LEX − in the syntax or lexically. We now postulate that a phrase is LEX − if and only if one of the daughters is LEX −, and that non-head-daughters of LEX − phrases are LEX −. This suffices to predict verbal complex formation, without requiring a special phrasal type like Müller (2008)'s *head-cluster-structure*. To account for linearization, we cross-classify *head-comp-phrase* with *liberating* and *compacting* and state that a *head-comp-phrase* is liberating if and only if it is LEX −.

TAM, Clause Type markers, and Subjects As we have noted above, tonology suggests that TAM markers form a constituent with the verbal complex. Thus, we analyze them as heads selecting a LEX + verb projection and taking over its argument requirements (15 left). As TAM markers are not part of the verbal complex, they are LEX −. We leave open how multiple TAM marking is analyzed, possibly, there is only one TAM marker which has multiple domain objects. Subjects are realized in a *spr-head-phrase* (15 right), which is also liberating. This phrase is used both for lexical and for clitical subjects.

(15)
$$\begin{bmatrix} \textsc{spr} & \boxed{1} \\ \textsc{comps} & \left\langle V \begin{bmatrix} \textsc{lex} & + \\ \textsc{spr} & \boxed{1} \\ \textsc{comps} & \boxed{2} \end{bmatrix} \right\rangle \oplus \boxed{2} \\ \textsc{lex} & - \end{bmatrix} \quad \begin{bmatrix} \textit{spr-head-phrase} & \\ \textsc{synsem|loc|cat|comps} & \langle\rangle \\ \textsc{head-dtr|s|l|cat|spr} & \left\langle \boxed{1} \right\rangle \\ \textsc{non-head-dtrs} & \left\langle \left[\textsc{synsem} \; \boxed{2} \right] \right\rangle \end{bmatrix}$$

Clause type markers are markers in the sense of Pollard & Sag (1994), attaching to saturated clauses. Via constraints on the marking value of clauses, one may

easily express the generalizations that clause type markers are optional, cannot be iterated, and are only allowed in matrix clauses. Prosodic constraints account for the facts that they cannot appear clause-initially (Brugman, 2009, 241) and cannot occur between the prefield and the subject.

Thus, the constituent structure we assume for the Khoekhoe clause is fairly standard: there is a binary-branching VP, in which complements and adjuncts may be realized in any order, TAM markers behave like raising verbs, and subjects and markers attach at a higher level. Semantic composition can proceed by the usual mechanisms and no extra stipulations are needed for clauses with fronting. The structure is essentially what an HPSG adaptation of the structure assumed by the GB/Minimalist analyses for clauses without fronting could look like. The major difference between the analyses is the mechanism used for fronting: While the Minimalist analyses assume that there is movement, we assume that clauses with fronting only differ in linearization, but not in constituent structure.

Example Analysis I Consider the basic examples in (2a–4) again. If we simplify the sentences by removing the NP complement for a moment, we arrive at (16a–d). For each sentence, it is also (at least marginally) possible to put *go* to the end (e.g., 16e).

(16)

	Prefield			*Middlefield*			
a. (tsî̂)		=s	ge	ǁari	go	maa=te	
and		=3fs	DECL	yesterday	TAM	give=1S	
b. (tsî̂)	ǁari	=s	ge	go		maa=te	
c. (tsî̂)	maa=te	=s	ge	go	ǁari		
d. (tsî̂)		=s	ge	go	ǁari	maa=te	
e. (tsî̂)		=s	ge		ǁari	maa=te	go

'(And) she gave me (something) yesterday'

All these sentences have the same constituent structure, which is shown in Figure 1. The VP, represented by an *head-adj-phrase*, consists of a verbal complex, a TAM marker, and an adjunct. The clause type marker and the enclitic subject pronoun attach at higher levels.

If the VP contains multiple dependents, the question arises in which order they attach. We will show below that they can be assumed to attach in the order in which they appear on the surface.

4.2 Constraining Linearization

The DOM list of the clause will contain all elements that can be permuted: the verbal complex, middlefield and prefield constituents, subject, and TAM markers.

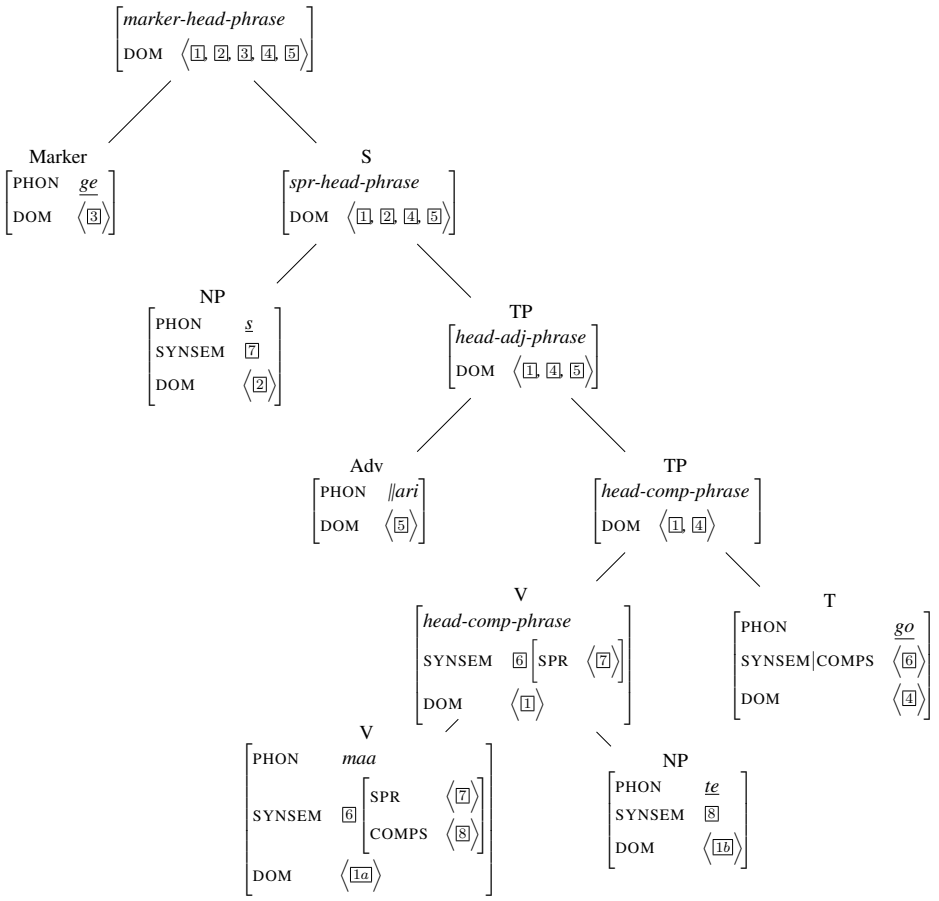

Figure 1: Analysis of (16). Clitics and monomoraic words are underlined.

Linear precedence rules will now restrict their relative ordering. We add a boolean-valued feature FRONTED appropriate for *dom-obj* that has the value + if and only if the sign is fronted. The members of the prefield and the subject count as fronted, while the elements following the subject are not fronted.

We can now formalize our generalization that, in the absence of monomoraic words, or fronting of the verbal complex, phrases are head-final. The following constraint appears to model this intuition:

(17) Head-Final Constraint (Preliminary Version)

$$\left[\begin{array}{ll} \text{HEAD-DTR} & \boxed{2}\left[\begin{array}{ll}\text{FRONTED} & -\\ \text{PHON} & \textit{multimoraic}\end{array}\right] \\ \text{NON-HEAD-DTRS} & \left\langle \boxed{1}\left[\text{PHON} \quad \textit{multimoraic}\right]\right\rangle \end{array}\right] \rightarrow \boxed{1} < \boxed{2}$$

It states that, if the head is not fronted and both the head and the dependent are multimoraic, a phrase is head-final. However, the constraint prescribes a relative ordering of the daughters, while it should be possible to shuffle the domains of the daughters. We therefore need a more general statement which quantifies over all domain elements of the daughters:

(18) Head-Final Constraint (Final Version)

$$\left(\begin{array}{l}\left[\begin{array}{ll}\textsc{hd-dtr} & \left[\textsc{dom} \left\langle \ldots\ \boxed{2}\left[\begin{array}{ll}\textsc{fronted} & \boxed{3}\\ \textsc{phon} & \textit{multimoraic}\end{array}\right]\ldots\right\rangle\right]\\ \textsc{n-hd-dtrs} & \left\langle \ldots\left[\textsc{dom}\left\langle \ldots\ \boxed{1}\left[\begin{array}{ll}\textsc{fronted} & \boxed{4}\\ \textsc{phon} & \textit{multimoraic}\end{array}\right]\ldots\right\rangle\right]\ldots\right\rangle\\ \textsc{dom} & \boxed{5}\end{array}\right]\\ \wedge\ \boxed{3}\dot{\rightarrow}\boxed{4}\end{array}\right)$$
$$\rightarrow \texttt{precede}(\boxed{5},\boxed{1},\boxed{2})$$

where `precede`(5,1,2) is true if and only if all occurences of 1 precede all occurences of 2 in the domain structure in 5. Here, $\dot{\rightarrow}$ is the usual implication relation between the *boolean* truth values $-$ and $+$: $\boxed{1} \dot{\rightarrow} \boxed{2}$ is true if and only if either $\boxed{1} = -$, or $\boxed{1} = \boxed{2} = +$. Thus, this constraint is more general in a second respect: it also applies when *both* selected domain elements are fronted. This becomes important in multiple fronting, where the verb, if it is fronted, is the last fronted element.

If both elements are non-fronted and multimoraic, the constraint says that the element from the head daughter precedes the element from the non-head daughter. In an utterance in which all phrases are headed and unary or binary, in which all words are multimoraic, and in which there is no fronting, this constraint enforces that all constituents are continuous and head-final, since the elements of the non-head daughter's domain will always precede those of the head daughter.

The constraint predicts that a non-fronted verbal complex follows the middlefield and that multimoraic TAM markers follow the VP. It also predicts that lexical subjects are incompatible with a nonempty prefield (6): Since all lexical roots are multimoraic in Khoekhoe (Brugman, 2009), neither a non-clitical subject nor an element of the prefield other than the clause-type marker can be mono-moraic. Thus, by (18), the domain object of the subject will precede all domain objects contributed by the VP.

If there are monomoraic or fronted elements, their linearization is not constrained by (18) and discontinuities may arise. The linearization of these elements is governed by (19):

(19) a. $[\text{FRONTED} +] < [\text{FRONTED} -]$

b. In a *spr-head-phrase*: $[\text{FRONTED} +] \leq$ Specifier

c. $[\text{HEAD } \textit{marker}] \leq [\text{FRONTED} -]$

d. $[\text{HEAD } \textit{tam-marker}] \rightarrow [\text{FRONTED} -]$

Fronted elements always precede non-fronted elements (19a), and the subject is the last fronted element (19b). This ensures that the prefield does not extend to the right of the subject clitic. The clause type marker precedes all (other) non-fronted elements (19c), but its FRONTED value and its position relative to the fronted elements are not constrained, allowing it to appear within the prefield or after the subject. Note that we use $\leq$ rather than $<$, as the specifier itself is also fronted. TAM markers are always non-fronted (19d). Monomoraic TAM markers may appear in any position that allows non-fronted elements, i.e. in the middlefield and after the verbal complex. Multi-moraic TAM markers are forced by (18) to appear after the middlefield and the verbal complex (9).

Constraining Multiple Fronting There is an additional constraint stipulating that at most one element may be fronted in declaratives (5), which may be stated as follows:

(20) $$\begin{bmatrix} \ldots\text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{CLAUSE-TYPE} \quad \textit{decl} \end{bmatrix} \\ \text{DOM} & \left\langle \textit{dom-obj} \right\rangle \oplus \boxed{1} \end{bmatrix} \rightarrow \boxed{1} \begin{pmatrix} \text{list}\left(\begin{bmatrix} \text{PHON} & \textit{prosodic-non-word} \end{bmatrix}\right) \\ \oplus \ \text{list}\left(\begin{bmatrix} \text{FRONTED} & - \end{bmatrix}\right) \end{pmatrix}$$

Together with (18), it excludes clauses such as (5). It might be considered preferable if this generalization somehow arose from the fronting mechanism. But the fact that it holds only for declaratives and that there are no other known differences between fronting in declaratives and interrogatives/hortatives suggests that this should not be expected. Rather, multiple fronting can be understood simply as one of the ways the language distinguishes non-declarative clauses from declarative ones, for which there is no general mechanism such as a question marker or inverted word order.

Example Analysis I Let us now examine how these linearization constraints work together to produce the linearizations in (16) for the constituent structure in Figure 1. By (19), the subject is FRONTED $+$ and the TAM marker is FRONTED $-$. By (19a), the subject therefore precedes the TAM marker. The linearization of the remainder depends on which elements are FRONTED $+$. If neither the verbal complex nor the adverb is FRONTED $+$, they both follow the subject. As they are multimoraic, constraint (18) applies to the *head-adj-phrase* and forces the adjunct to appear in front of the verbal complex. On the other hand, the position of the

monomoraic TAM marker is not restricted by this constraint, and (16c–e) are licensed. By (19c), the clause type marker must appear to the left of these elements. As the clause type marker cannot appear to the left of the subject by prosodic constraints, it can only be placed immediately after the subject.

If, say, the verbal complex is FRONTED +, it precedes all other elements by (19a). Constraint (18) does not apply, and the head-initial ordering is licensed. Similarly, if the adverb is FRONTED +, orderings such as the one in (16b) result.

Example Analysis II: Multiple Fronting In the previous example, the verb had only one non-clitical dependent. Let us now see what happens if there are several dependents, as in the following example with a ditransitive and multiple fronting:

(21) taras-a kha maa=ts go ǂkhanis-a ǁari?
woman-A INTERROG give=2MS TAM book-A yesterday
'did you give the woman a book yesterday?'

The VP contains a verb, an adverb, and two argument NPs. The verb and one complement are fronted. The clause type marker and the enclitic subject pronoun attach at higher levels, but again are linearized within the VP. This sentence has the syntactic structure in Figure 2. The difference to Figure 1 is that the VP is more complex and contains two complements in addition to an adjunct. The order in which these attach to the verb directly mirrors the order in which they are linearized. This is also the only possible analysis. To see this, suppose that, say, the complement *ǂkhanisa* and the adjunct *ǁari* were interchanged in the tree. Then there is an *head-adj-phrase* combining the adjunct (ǁari) with the constituent formed by the complement (ǂkhanisa), the TAM marker, and the verb (maa), where the latter daughter is the head. Now if 1 is the complement and 2 the adjunct, constraint (18) applies to the *head-adj-phrase* and forces the adjunct to precede the complement, different from the ordering in (21). As the constraint in (18) targets all domain elements of the head, it also applies if the verb is fronted, which means that there is only one analysis even if the verb is fronted. Thus, the use of word order domains does not lead to spurious ambiguities such as those that have been noted for domain-based analyses of German word order (Müller, 2008, 152).

5 Discussion

In Section 3.3, we claimed that a linearization-based analysis provides a simple description of the data which directly motivates patterns that are not predicted by the previous movement-based accounts. The phrasal types and lexical entries are very simple and essentially identical to ones that have been previously proposed for other languages. The mechanism of word-order domains has been successful crosslinguistically, and we only used the basic distinction between liberating

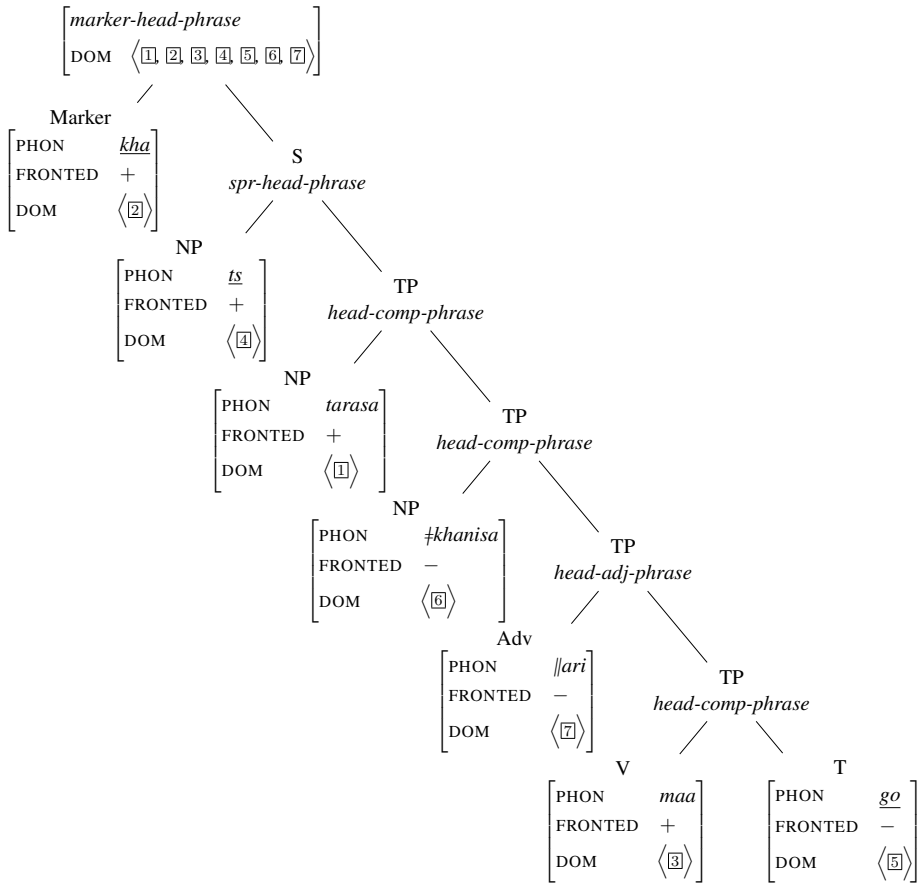

Figure 2: Analysis of (21). The domain objects are numbered by their linear order. Note that, for multimoraic non-heads, linear order coincides with the order of realization in the tree.

and compacting phrases. The only complex language-specific constraints are the Head-Final Principle (18), which we motivated independently in Section 3.3, and the constraint on multiple fronting (20), which is simply one of the mechanisms Khoekhoe uses for marking clause type.

It can also be noted that, assuming that Khoekhoe words receive sandhi tone whenever they are a noninitial daughter and stipulating Haacke's rule of 'Relative Retention', our analysis of constituent structure correctly predicts the tone sandhi data in Haacke (1999, Chapter 4). Some of the additional data in Brugman (2009) is not explained, but it appears likely that some additional rules need to be stipulated in any case. The previous analyses did not consider tone and do not have a sufficient coverage to test them on all of Haacke's data, but it is in principle also compatible with extraction-based analyses.

Our analysis is somewhat similar to the cross-linguistic account of V2 proposed by Wetta (2011), who defines an attribute LIN appropriate for *sign* with values *flexible* and *fixed*. *Fixed* elements are those whose position is determined constructionally, such as the verb in V2 clauses. Linearization constraints such as those acting on a 'middlefield' will typically affect flexible elements without interfering with fixed elements. This feature can be identified with our FRONTED attribute. As Khoekhoe does not have V2, our linearization constraints are quite different from Wetta's, but our analysis confirms the applicability of his general approach.

Alternatives without Domains or Extraction One might analyze fronting via a local extraction mechanism that is applicable to both the head of the clause and its dependents, uniting local extraction and head movement. The difficulty with this is that such an account requires a mechanism that works for both heads and their dependents, which to our knowledge has not yet been proposed in the HPSG literature. It seems preferable to use a cross-linguistically well-motivated mechanism rather than a mechanism designed for a single language.

One might also assume that fronting arises from the ability of verbs to realize their dependents in different orders, as has been suggested in HPSG for German verb placement (Crysmann, 2004). Such an approach faces difficulties when the verb apears in the prefield (4c), as the relative order of subject, clause type marker, and middlefield remains invariant, while it would be excepted to be reversed. Note that the fact that TAM markers are probably heads is irrelevant, as the availability of fronting of the verbal complex does not depend on their presence. Furthermore, there are word order rules applying to the middlefield that appear to be unaffected by fronting, in particular the discontinuous TAM markers. In such an analysis, these rules would have to be stipulated either via linearization domains, in which case both complex variation in constituent structure and linearization domains are required for the analysis of the same phenomenon, or locally by some otherwise unmotivated book-keeping mechanism.

A further alternative is a completely flat structure, in which fronting can be described as permutation of sister nodes, as in the analyses assumed by Haacke (1978). Besides the verb, complements and adjuncts, this structure would also contain the subject and the clause type marker. However, such an analysis requires significant formal machinery to model the possible configurations. Such an analysis can be seen as a notational variant of our analysis in which the realization of dependents is regulated not by simple rules operating on larger binary-branching structures, but by complex rules operating on simple phrase structures.

The Postfield It is possible for one or more adjuncts or complements to occur after the verbal complex (Hagman, 1977, 113). This may be taken as evidence that the verbal complex is part of the middlefield, which would then extend to the right, and that verb placement is free. This appears to agree well with the fact that TAM

markers may appear postverbally. However, it seems that postverbal elements are always instances of extraposition. First, it is impossible for a postverbal element to intervene between the verb and a postverbal TAM marker. Second, Khoekhoe has an unbounded dependency extraposition pattern for arguments where there is a pronoun at the canonical position (cf. Haacke (1992) for a comprehensive discussion). Such extraposed elements may intervene between the verbal complex and a postverbal argument (22a), but may not occur in front of the verbal complex (22b).

(22) a. taras ge go maa=gu$_i$ [‖naa khoega]$_i$ [Petrub go xoa
woman DECL TAM give=them those men-A Petrus TAM write
ǂkhanisa]
book-A
'the woman gave those men the book Petrus wrote'

b. * taras ge go [‖naa khoega]$_i$ maa=gu$_i$ [Petrub go xoa
woman DECL TAM those men-A give=them Petrus TAM write
ǂkhanisa]
book-A
'the woman gave those men the book Petrus wrote'

Thus, we assume that the rightmost position for TAM markers is the position immediately following the (non-fronted) verbal complex, while arguments and adjuncts may be extraposed with or without a pronoun in the canonical position.

6 Conclusion

Khoekhoe allows local fronting of both predicates and complements/adjuncts. We showed that this can be analyzed as a discontinuity of the VP, which is interrupted by monomoraic words attaching at a higher level. More generally, we showed that monomoraic words are exempt from the general head-final order of Khoekhoe and argued that they can give rise to discontinuous constituents. We provided a formal HPSG analysis, showing how Khoekhoe word order variation can be analyzed without empty elements based on mechanisms that have previously been proposed for other languages. The analysis has a significantly wider empirical scope than the previous Minimalist analyses of Khoekhoe clause structure.

References

den Besten, Hans. 2002. Khoekhoe syntax and its implications for L2 acquisition of Dutch and Afrikaans. *Journal of Germanic Linguistics* 14.1. 3 – 56.

Brugman, Johanna Christina. 2009. *Segments, tones and distribution in Khoekhoe prosody*: Cornell University dissertation.

Crysmann, Berthold. 2004. On the efficient implementation of German verb placement in HPSG. In *Proceedings of RANLP 2003*, 112 – 116.

Donohue, Cathryn & Ivan A. Sag. 1999. Domains in Warlpiri. In *Sixth international conference on HPSG – Abstracts*, 101–106. Edinburgh.

Haacke, Wilfrid H. G. 1976. *A Nama grammar: the noun-phrase*. University of Cape Town MA thesis.

Haacke, Wilfrid H. G. 1978. *Subject deposition in Nama*. Colchester University of Essex MA thesis.

Haacke, Wilfrid H. G. 1992. Dislocated noun phrases in Khoekhoe (Nama/Damara): further evidence for the sentential hypothesis. *Afrikanistische Arbeitspapiere* 29. 149–162.

Haacke, Wilfrid H. G. 1999. *The tonology of Khoekhoe (Nama/Damara)* Quellen zur Khoisan-Forschung. Köppe.

Haacke, Wilfrid H. G. 2013. Syntax: Namibian Khoekkhoe, !Gora. In Rainer Vossen (ed.), *The Khoesan languages*, 325–346. Routledge.

Haacke, Wilfrid H. G. & Eliphas Eiseb. 2002. *A Khoekhoegowab dictionary with an English-Khoekhoegowab index*. Gamsberg Macmillan.

Hagman, R.S. 1977. *Nama Hottentot grammar*, vol. 15. Indiana University Bloomington.

Huybregts, Riny. 2003. A PF-interface condition on syntactic computation. In Jan Koster & Henk van Riemsdijk (eds.), *Germania et alia: A linguistic webschrift for Hans den Besten*, . `http://www.let.rug.nl/koster/DenBesten/contents.htm`.

Kasper, Robert T. 1994. Adjuncts in the mittelfeld. In John Nerbonne, Klaus Netter & Carl J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 39–69. Stanford University: CSLI Publications.

Kathol, Andreas & Carl Pollard. 1995. Extraposition via complex domain formation. In *Proceedings of the 33rd annual meeting of the association for computational linguistics*, 174 – 180.

Klein, Harriet E. M. 1976. Tense and aspect in the Damara verbal system. *African Studies* 35. 207–227.

Krönlein, Johann Georg. 1889. *Wortschatz der Khoi-Khoin*. Berlin: Deutsche Kolonialgesellschaft.

Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche* (Linguistische Arbeiten 394). Tübingen: Max Niemeyer Verlag.

Müller, Stefan. 2002a. *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford: CSLI Publications.

Müller, Stefan. 2002b. Multiple frontings in German. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Winter (eds.), *Proceedings of formal grammar 2002*, 113–124. Trento.

Müller, Stefan. 2008. *Head-Driven Phrase Structure Grammar: Eine Einführung* (Stauffenburg Einführungen 17). Tübingen: Stauffenburg Verlag 2nd edn.

Müller, Stefan. 2013. German sentence structure: An analysis with special consideration of so-called multiple frontings. Manuscript, `http://hpsg.fu-berlin.de/~stefan/Pub/gs.html`.

Müller, Stefan. To Appear. HPSG – A Synopsis. In Artemis Alexiadou & Tibor Kiss (eds.), *Syntax – Ein internationales Handbuch zeitgenössischer Forschung* Handbücher zur Sprach- und Kommunikationswissenschaft, Berlin: Walter de Gruyter Verlag 2nd edn.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Reape, Mike. 1994. Domain union and word order variation in German. In John Nerbonne, Klaus Netter & Carl J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 151–197. Stanford University: CSLI Publications.

|Uriseb, A. 1993. *‖Hûisa !nubu ‖gaeǂhôadi*. Windhoek: Gamsberg Macmillan.

Washburn, P. 2001. A Minimalist approach to Khoekhoe declaratives. *Cornell Working Papers in Linguistics* 18. 28–56.

Wetta, Andrew C. 2011. A construction-based cross-linguistic analysis of V2 word order. In Stefan Müller (ed.), *The Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, 248–268. Stanford: CSLI Publications.

Witzlack-Makarevich, Alena. 2006. *Aspects of Information Structure in Richtersveld Nama*. Universität Leipzig MA thesis.

Zwicky, Arnold M. & Geoffry K. Pullum. 1983. Clitization vs. Inflection: English n't. *Language* (59). 502–513.