Word Order Variation in Khoekhoe

Michael Hahn University of Tübingen

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. Khoekhoe has been claimed repeatedly to have a clause-second position (Washburn 2001, den Besten 2002, Brugman 2009) and to have topological fields similar to German (Witzlack-Makarevich 2006). However, previous generative analyses in GB/Minimalism (Washburn 2001, Den Besten 2002, Huybregts ms.) are based on descriptive grammars and make empirical assumptions that turn out to be problematic in the light of corpus data and elicited data. Based on new data, we show that the phenomenon can be analyzed by allowing the Khoekhoe VP to be discontinuous, with monomoraic words intervening, while multimoraic words and phrases are subject to strict head-final ordering. The analysis is similar to the cross-linguistic account of V2 proposed by Wetta (2011), confirming the applicability of the general approach to languages with positionally stipulated word order.

Basic Data

Khoekhoe is a head-final SOV language (1a). In matrix clauses, one of the clause type markers *ge* (*declarative*), *kha* (*interrogative*), *ko/kom* (*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 (1 b-d). By analogy to German, we will refer to the collection of these elements as the *Middlefield*, following Witzlack-Makarevich (2007):

(1) a.	[khoexa-khoeb	âts]	(ge)	llari	‡khanis-a	go	ma=te
	friend-Змs	Poss-2ms	DECL	yesterday	book.3fs-case	TAM	give=me
b.	khoexa-khoeb	âts	(ge)	‡ khanisa	llari	go	ma=te
c.	khoexa-khoeb	âts	(ge)	llari	go	‡khanisa	ma=te
d.	khoexa-khoeb	âts	(ge)	go	llari	‡khanisa	ma=te
'you	r friend gave me	a book yest	erday'	_			

Bimoraic TAM markers (tide, hâ, i) follow the verbal complex. Complement NPs are realized as full noun phrases with the case marker -*a* (glossed CASE) or as a pronominal clitic in the verbal complex. A subject may be realized by a full NP or an enclitic **P**erson-**G**ender-**N**umber (PGN) marker appearing in the ordinary subject position (2). In the latter case, a coreferent case-marked NP that behaves much like a complement may appear in the middlefield. We assume that this NP is in fact an ordinary complement introduced by a lexical rule and will ignore it in the rest of the paper.

(2)	tsî=s	ge	(tarasa)	llari	‡khanisa	go	maa=te
	and=3 _{FS}	DECL	woman-3fs-case	yesterday	book-3fs-A	TAM	give=me
	'and she / the woman gave the man a book'						

If the subject is realized by a PGN marker only, a complement (3a), an adjunct (3b), or the verbal complex itself (3c) may precede it. In declaratives, there may be at most one such element (3d). The subject position cannot be occupied by a lexical NP in this case (3e):

		[Prefield]			[Middlefiel	d]	Verbal Complex
(3)	a.	‡ khanisa	$=_S$	ge	go	llari			maa=te
		book-case	3 _{FS}	DECL	TAM	yesterday			give=me
	b.	llari	$=_S$	ge	go		‡ khanisa		maa=te
	c.	maa=te	$=_S$	ge	go	llari	‡ khanisa		
	d. *	‡khanisa maa=te	$=_S$	ge	go	llari	‡ khanisa		
	e. *	+ khanisa	taras	ge	go	llari			maa=te
		book-case	woman	DECL	TAM	yesterday			give=me

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*. Note that, unlike Germanic V2, fronting into the 'prefield' is always local in Khoekhoe. It is also not possible to only front parts of the verbal complex:

```
(4) a. !gû lloa=b a
go not.be.able=3ms TAM
b. * !gû=b a lloa
c. * lloa=b a !gû
'he cannot walk' (after Krönlein, 1889)
```

While the prefield may contain at most one element in declaratives, it may contain several elements in interrogative and hortative clauses. While these elements are usually one or more dependents followed by the verb (5a), this is not the only possibility. In fact, any sequence of elements that could occur alone in the prefield seems to be allowed, cf. (5b-d).

(5)	a.	tarasa	‡khanisa	maa=b	go	llari?			
		woman-case	book-case	give=3 _{MS}	TAM	yesterday			
	b.	tarasa	maa	‡khanisa=b	go	llari?			
	c.	maa	tarasa	‡khanisa=b	go	llari?			
	d.	tarasa	‡khanisa=b	go	llari	maa			
	e.	‡khanisa	tarasa=b	go	llari	maa			
	'did he give the woman a book yesterday?'								

The clause type marker *kha* may appear between any two prefield elements (6a-c), while *kha* may never appear in the middlefield (6d). *Kha* cannot be iterated (6e).

(6)	a.	tarasa kha	maa		‡khanisa=b		go		llari?
	b.	tarasa	maa	kha	‡khanisa=b		go		llari?
	c.	tarasa	maa		‡khanisa=b	kha	go		llari?
	d.	* tarasa	maa		‡khanisa=b		go	kha	llari?
	e.	* tarasa kha	maa	kha	‡khanisa=b		go		llari? (etc.)

Other Fronting 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. 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. There is no wheextraction in Khoekhoe.

Analysis

Khoekhoe word order and clause structure has been studied in GB/Minimalism by Washburn (2001), den Besten (2002), and Huybregts (ms.). They all analyze the prefield via movement, and assume that the clause type marker (Washburn, 2001; den Besten, 2002) or the subject PGN marker (Huybregts, m.s.) occupy the highest (overtly filled) head position. Unfortunately, these studies seem to rely entirely on descriptive grammars, making incorrect empirical assumptions. For instance, the presentation in Hagman (1977)'s grammar suggests that TAM markers are always adjacent to the verbal complex unless it is fronted. Under 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 (3d). However, the empirical claim is falsified by (1 c-d)², and hence there is no motivation for the ungrammaticality of (3d), nor does the analysis account for the grammaticality of (5). Similar problems occur in the other analyses.

We will now develop an HPSG analysis that avoids these empirical problems. We will first explore approaches making use of variation in constituent structure similar to previous GB/Minimalist analyses, and argue that a linearization-based analysis is more adequate.

Word Order Variation as Variation in Constituent Structure?

A first approach is to assume that the word order variation in (1) and (3) arises from the ability of verbs to realize their dependents in different orders, subject to certain constraints ruling out clauses like (3d). While it is conceivable that one complement is realized at a higher position than the subject in (3 a-b), the situation becomes more difficult if the verb apears in the prefield (3c). Not only does the relative order of subject, clause type marker and middlefield remain invariant, there are also word order rules applying to the middlefield that are unaffected by fronting. There are discontinuous TAM markers *go...a*, *ge...a*, whose order remains invariant under fronting of the verb. Furthermore, the clause type marker is always realized after the subject if it is

Hagman (1977, 144) only mentions the fronting of complements together with the verb in clauses with (in our terminology) a clitic-left-dislocated subject, but the possibility of multiple fronting without a fronted subject is confirmed by corpus data, cf. [lawe go haa haab-a]_{NP} mû=ts go xare? just TAM come horse-CASE see=2MS TAM PART 'did you see the horse that just came?' (Krönlein, 1889). Multiple fronting without a fronted verb is rare and not mentioned by the descriptive grammars, but it is attested in corpus data.

² Other examples are given by Haacke (1999), Witzlack-Makarevich (2006), and Brugman (2009, 244). There is also a fairly significant amount of corpus evidence, cf. tsî=ta ge tiita ra !âs-i lloo. and=1s DECL I TAM hunger-ABLATIVE die 'and I am dying of hunger.' (Krönlein, 1889)

realized after the prefield.

It seems more reasonable to 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, with the syntactic structure and the position of the remainder remaining invariant.

Under such an approach, we might assume that the highest head is not the verb, but an element that resists preposing, i.e. a TAM marker or the clause type marker. Thus the prefield may be analyzed as arising from a uniform mechanism that preposes complements and adjuncts, for instance the SLASH feature. This is essentially the perspective taken by Washburn (2001), who assumes that the TAM marker occupies T and that all non-clitical dependents have to move outside of vP, so that preposing can be analyzed uniformly as the movement of phrases to Spec-FocP.

It is unexpected under such an analysis that fronting into the prefield seems to always be local in Khoekhoe, while the language allows unbounded extraction with gaps in relative clauses. Furthermore, clauses that lack overt TAM marking show the same word order possibilities, including fronting of the verbal complex:

(7) [‡nû tama]=s ge lôas-a neepa sit NEG=3fs DECL daughter-CASE here 'the daughter has not sat here' (Hagman 1977, 112)

Under such an analysis, a uniform analysis of the prefield can only be achieved by stipulating empty TAM elements or by some other way of mimicking the effect of an empty head. Similarly, an empty head is required if fronting is motivated by treating the clause type marker as the head, as it is never obligatory, or if the subject PGN marker is analyzed as an inflected complementizer (Huybregts ms.), as it is optional in clauses like (8), where the nonreferential subject need not be expressed:

(8) [\pm tama](='i) kom [llg\hat{a}us ai]=0 right NEG=3NS ASSERT home in-ASSERT 'something is not right at home'

As an alternative, one might analyze clauses with a filled prefield via a local extraction mechanism that is applicable to both heads and dependents, i.e. maximal projections. 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. We will now show how Khoekhoe word order can be analyzed without empty elements and based on mechanisms that have previously been proposed for other languages.

Linearization-based Analysis

For several languages with free word order that does not seem to match the expected constituent structure, analyses using discontinuous constituents have been developed in HPSG (e.g., Reape 1994, Kathol 2000, Donohue and Sag 1999). Thus, we might analyze Khoekhoe fronting in terms of linearization, rather than constituent structure. In fact, as there seem to be no concomitant morphosyntactic effects,³ fronting seems to be purely a matter of word order and information structure descriptively. Language-internally, there appear to be reasons to link fronting to different tree structures – except that word order is generally assumed to be tightly linked to constituent structure.

Brugman (2009) observes that some of the constraints on the linearization of PGN markers and clause type markers follows from the fact that they are consist of at most one mora. According to her analysis, mono-moraic words are not prosodic words in Khoekhoe, and hence cannot appear in a clause-initial position. This suggests that PGN markers and clause type markers are second-position clitics, whose linearization might be amenable to a treatment combining syntactic and prosodic elements. In fact, monomoraic words more generally tend to show unexpected word order patterns in Khoekhoe, when compared to their multimoraic counterparts. With the exception of verbal complex fronting (3c) and extraposition, the latter seem to strictly follow head-final word order. For instance, prosodically autonomous complements generally precede the head, while object clitics follow the verb. There is a similar difference in the position of TAM markers. While monomoraic TAM markers (ra, ge, go, nî, a, ga) and their combinations (ge-re, go-ro, nî-ra, ga-ra) are generally placed in the middlefield, and may also follow the verbal complex (9), bimoraic TAM markers (hâ /hââ/, tide, i /ii/) must always follow the verbal complex (9 b-c).

- ge (9)namas a. maa ra tarasa Nama(f.) DECL give TAM woman 'the Nama (woman) is giving (something) to the woman' b. namas ge tarasa maa tide Nama(f.) DECL woman give TAM+NEG 'the Nama (woman) will not give (anything) to the woman' c. * namas tarasa tide maa
- 3 Hagman's (1977, 112) generalization that there is obligatory deletion of the TAM marker $h\hat{a}$ when the verb is fronted is incorrect, as it is possible to have $h\hat{a}$ with a fronted verb. This is confirmed by corpus data, e.g. !kûi=s hâ tsâusa be.pregnant=3FS TAM heifer-CASE 'is the heifer pregnant?' (Krönlein, 1889). It is also possible to drop $h\hat{a}$ without verb fronting.

We suggest that all TAM markers are heads that require a VP complement. Monomoraic words, not being prosodic words, are exempt from the general head-final word order. Thus, obligatory posterbal position of bimoraic TAM markers, which would require additional stipulations if TAM markers were treated as complements or via a construction, follows automatically. Monomoraic words may give rise to discontinuous constituents, as in (1 a-c), where the TAM marker is placed within the VP, i.e., its complement.

The same treatment may be applied to the unexpected word order properties of subject PGN markers and clause type markers: they attach to a satured projection of a verb or a TAM marker, but may be realized within the VP, which thus becomes discontinuous. This treatment is compatible with the fact that the placement of these elements is sensitive mainly to the *number* of elements in the prefield, but insensitive to their syntactic categories.

We conclude that Khoekhoe is a strict head-final language, but that fronted constituents and monomoraic words are exempt from this constraint, and that their linearization behavior may give rise to discontinuous constituents.

Formalization

In our informal discussion, we suggested that the Khoekhoe VP is often discontinuous. This idea can be formalized in domain-based HPSG (Reape 1994). Every *sign* is assigned a domain encoded in the DOM list, whose elements are again of type *sign*. The phonology of a phrase is computed not from the phonologies of the daughters, but from the phonologies of the domain elements:

(10)
$$sign \to \begin{bmatrix} PHON & \boxed{1} \oplus ... \oplus \boxed{n} \\ DOM & \left\langle \begin{bmatrix} PHON & \boxed{1} \end{bmatrix}, ..., \begin{bmatrix} PHON & \boxed{m} \end{bmatrix} \right\rangle$$

Phrasal constructions differ as to how their domain is computed from the domains of their daughters. *Liberating* constructions take all domain elements of the daughters into their domain, while *compacting* constructions fuse the daughter's domains into a single domain element (Kathol and Pollard, 1995; Donohue and Sag, 1999). By using liberating constructions, domain structure can be dissociated from constituent structure, and it becomes possible to model the idea of discontinuous constituents and to impose linearization constraints on elements that may be far apart in constituent structure without tree-traversal.

In Khoekhoe, the verbal complex, noun phrases, and postpositional phrases are *compacting*, while higher projections of the verbal complex are *liberating*, allowing elements attaching on higher levels to intervene. The VP could either be flat, have a binary branching structure, or an intermediate structure where complements are realized together, while adjuncts attach via a binary rule, and can then be 'mixed' with the complements via discontinuous constituents. We have not been able to find empirical evidence that any intermediate constituents between the VP and the verbal complex exist in Khoekhoe. Unless such evidence appears, an internal constituent structure would thus probably not be much more than a formal device for doing the syntactic and semantic computations that would otherwise be expressed by a relational constraint. We will assume a flat VP, which is licensed by the *head-comps-adjs-phrase* type (Kasper 1994), subject to the following constraint:

The boolean-valued feature LEX is used to enforce verbal complex formation and obligatory coherence (Müller 2002, 87). As a side effect, it prevents multiple applications of the *head-comps-adjs* rule, like Abeillé and Godard's (2002) WEIGHT feature. Assuming that PGN clitics are specified as LEX +, the constraint also ensures that only full NPs are allowed in complement positions.

The subject is realized in a *subject-head-phrase*, which is also *liberating* in Khoekhoe. Clause type markers are analyzed as markers in the sense of Pollard & Sag (1994), attaching to saturated clauses. Via appropriate constraints on the MARKING value of clauses, this allows for a straightforward motivation of the fact that they are optional, cannot be iterated (5e), and are only allowed in matrix clauses. The fact that they cannot appear clause-initially is predicted by a general constraint requiring the left edge of a phonological phrase to be the left edge of a prosodic word, as the clause type markers are not prosodic words (Brugman, 2009, 241). Similarly, they cannot occur between the prefield and the subject because the subject PGN clitic can only attach to prosodic words. These prosodic constraints will be formalized in the full paper.

Constraining Linearization

Since the projections of the verbal complec are *liberating*, the subject, the clause type marker, TAM markers, the verbal complex, and its dependents are all on the DOM list of the clause. The word order variation found in the Khoekhoe clause can be expressed by

linear precedence rules operating on this DOM list. We follow Wetta (2011), a cross-linguistic analysis of V2, in using an attribute LIN appropriate for *sign* with values *flexible* and *fixed*. Elements whose position is determined constructionally, such as the verb in V2 clauses, are *fixed*. Liberating phrases are assumed to be *flexible*. Linearization constraints will typically affect *flexible* elements without interfering with *fixed* elements. In Khoekhoe, members of the prefield and the subject are *fixed*, the clause type marker may have either value, and other constituents are *flexible*.

As shown above, Khoekhoe is a head-final language, but fronted constituents and monomoraic words are exempt from this constraint. Discontinuity is only available in the their presence. The following general principle formalizes this idea:

Let X be a phrase headed by a *flexible* multi-moraic element. Then the multi-moraic non-head daughters are inserted in the DOM list of X in front of the domain elements of the head daughter.

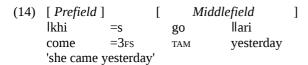
In essence, this states that, in the absence of monomoraic words, or fronting phrases are continuous and strictly head-final. It predicts that a non-fronted verbal complex follows the middlefield and that multi-moraic TAM markers follow the VP. It also predicts that lexical subjects are incompatible with a nonempty prefield (3e): Since all lexical roots are multimoraic in Khoekhoe (Brugman, 2009), neither a lexical subject nor a VP can be mono-moraic. Thus, by (12), the domain object of the subject will precede all domain objects contributed by the VP. On the other hand, the mono-moraic clause type markers may still intervene as in (1), even though they attach at a higher level.

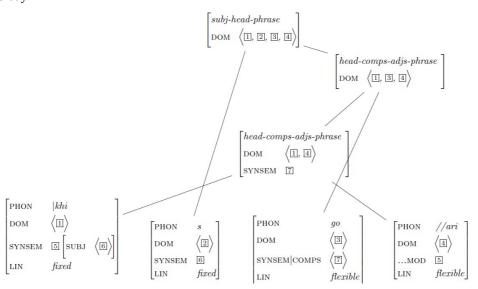
The positioning of *fixed* and monomoraic elements is left to other constraints, shown in (13). *Fixed* elements always precede *flexible* elements (13a). The subject PGN marker is the last *fixed* element (13b), ensuring that the prefield does not extend to the right of the subject PGN marker. The clause type marker precedes all other *flexible* elements (13c), but its position relative to the *fixed* elements is not constrained, allowing for the variability observed in (6). TAM markers are always *flexible* (13d). This is sufficient to predict the word order variation of TAM markers: Monomoraic TAM markers may appear in any position that allows flexible elements, i.e. in the middlefield and after the verbal complex. Multi-moraic TAM markers are forced by (12) to appear after the VP, i.e., after the middlefield and the verbal complex (9 b).

- (13) a. [LIN fixed] < [LIN flexible]
 - b. $[LIN fixed] \leq [HEAD pgn-marker]$
 - c. $[HEAD marker] \leq [LIN flexible]$
 - d. $[HEAD tam-marker] \rightarrow [LIN flexible]$

Example Analysis

The figure below shows an analysis of (14), a simplified version of (3c) with an intransitive verb and without clause type marker. The VP, represented by a discontinuous *head-comps-adjs-phrase*, consists of a *fixed* verb and an adjunct. The TAM marker and the enclitic subject pronoun attach at higher levels, but they are linearized within the VP.





Constraining Multiple Fronting

There is an additional constraint stipulating that at most one element may be fronted in declaratives, which may be stated as follows:

$$\begin{bmatrix} verb \\ CLAUSE-TYPE & decl \end{bmatrix} \longrightarrow \mathbb{I} \left(list \Big(\begin{bmatrix} PHON & prosodic-non-word \end{bmatrix} \right) \oplus list \Big(\begin{bmatrix} LIN & flexible \end{bmatrix} \right)$$

Together with (13 a-b), it excludes clauses such as (3d). It might be considered preferable if this generalization somehow arose from the fronting mechanism. However, 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.

An alternative analysis might directly encode topological fields in the domain structure, using the general formal mechanisms described by Penn (1999). Such an analysis might be considered to have greater explanatory adequacy, but it would require more complex formal machinery. It is also possible to realize clitics and fronted elements using a feature percolation mechanism and one or more unary projection rules whose head is not realized locally, thus eliminating discontinuous constituents from the analysis. However, we believe that a domain-based analysis captures the intuition of discontinuous constituents more directly.

References

Anne Abeillé and Daniele Godard. The syntactic structure of French auxiliaries. Language, 78(3):404 – 452, 2002. Hans den Besten. Khoekhoe syntax and its implications for L2 acquisition of Dutch and Afrikaans. Journal of Germanic Linguistics, 14.1:3 – 56, 2002.

Johanna Christina Brugman. Segments, Tones and Distribution in Khoekhoe Prosody. PhD thesis, Cornell, 2009.

Cathryn Donohue and Ivan A. Sag. Domains in Warlpiri. Sixth International Conference on HPSG – Abstracts, pages 101-106. Edinburgh, 1999.

Jonathan Ginzburg and Ivan A. Sag. Interrogative Investigations: the form, meaning, and use of English Interrogatives. CSLI Publications, Stanford, California, 2001.

Wilfrid H. G. Haacke. The Tonology of Khoekhoe (Nama/Damara). Köppe, 1999.

Roy S. Hagman. Nama Hottentot grammar. Indiana University Bloomington, 1977.

Riny Huybregts. A PF-interface condition on syntactic computation. http://odur.let.rug.nl/koster/DenBesten/Huybregts.pdf.

Robert T. Kasper. Adjuncts in the Mittelfeld. In John Nerbonne, Klaus Netter, and Carl J. Pollard, editors, German in Head-Driven Phrase Structure Grammar, pages 39 – 69. CSLI Publications, Stanford, 1994.

Andreas Kathol. Linear Syntax. Oxford University Press, Oxford, 2000.

Andreas Kathol and Carl Pollard. Extraposition via complex domain formation. In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, pp. 174–180. 1995

Johann Georg Krönlein. Wortschatz der Khoi-Khoin. Berlin, 1889.

Stefan Müller. Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche. Number 394 in Linguistische Arbeiten. Max Niemeyer Verlag, Tübingen, 1999.

Stefan Müller. Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German. CSLI Publications, Stanford, 2002.

Gerald Penn. An RSRL formalization of Serbo-Croatian second position clitic placement. In Valia Kordoni, editor, Tübingen Studies in Head-Driven Phrase Structure Grammar, pages 177 – 197, Tübingen, 1999.

Mike Reape. Domain union and word order variation in German. In John Nerbonne, Klaus Netter, and Carl J. Pollard, editors, German in Head-Driven Phrase Structure Grammars, pages 151 – 197. CSLI Publications, 1994.

Paul Washburn. A minimalist approach to Khoekhoe declaratives. Cornell Working Papers In Linguistics, 18:28 – 56, 2001. Andrew C. Wetta. A construction-based cross-linguistic analysis of V2 word order. In Stefan Müller, editor, The Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar, pages 248 – 268, Stanford, 2011.

Alena Witzlack-Makarevich. Aspects of information structure in Richtersveld Nama. Master's thesis, Leipzig, 2006.