The grammaticalisation of verb-auxiliary order in East African Bantu: from information structure to tense-aspect

Hannah Gibson
University of Essex

Abstract
Bantu languages employ a combination of simple and compound verb forms to encode tense-aspect-mood distinctions. Compound constructions typologically involve an auxiliary form followed by an inflected main verb. However, the six East African Bantu languages under examination in this paper exhibit a typologically and comparatively unusual construction in which the auxiliary appears after the verb. This paper presents a synchronic description of this word order and develops an account of its possible origins. It is proposed that the verb-auxiliary order originated from a verb-fronting construction which was used historically to convey predication focus. The account further corroborates the claim that the progressive aspect is an inherently focal category in Bantu and from a wider perspective, shows the interplay between the encoding of information structure and tense-aspect information.

1 Introduction

Information structural properties such as focus, and tense-aspect distinctions are often thought to belong to distinct domains of linguistic organisation: pragmatics and semantics, respectively. However, there is evidence from Bantu languages that a range of markers across the language family concurrently fulfil both pragmatic and semantic functions. This paper examines the interplay between the encoding of focus and tense-aspect information, through an examination of the grammaticalization of an unusual word order in a set of East African Bantu languages.

Bantu languages use a combination of morphological markers and auxiliary forms to encode a rich array of tense-aspect distinctions. A simple verb form is comprised of a single verb which may be inflected for tense and aspect through the use of affixes and the associated tonal pattern. Compound verbal constructions typically involve one or more auxiliary form followed by an inflected main verb (Nurse 2008:142; Gibson 2019).

A subset of East African Bantu languages exhibits an unusual order in which the main verb appears before the auxiliary. This construction is typologically unexpected since SVO languages are commonly associated with pre-verbal auxiliary placement cross-linguistically (Greenberg 1966:83, Dryer 1992:100). This order is also atypical in the comparative context of East African Bantu where auxiliary-verb structures dominate. The examples below from the Bantu languages Rangi spoken in Tanzania (1) and Kuria spoken in Kenya (2) are illustrative of the construction under examination where the main verb precedes the inflected auxiliary form (shown here in bold).[1,2]

[1] Glosses follow the Leipzig Glossing Rules with the following additions: 1, 2, 3, etc.= noun class; ANT=anterior; ASSOC=associative; AUG=augment, AUX=auxiliary; CJ=conjoint, CON=connective; CONS=consecutive; CONT=continuous; DJ=disjoint, DIR=directional; FPST=far past; FV=final vowel; HAB=habitual; IMPF=imperfective; ITR=iterative; NAR=narrative; NPST=non-past, OM=object marker; PAST1=recent past; PAST 2=distant past; PREP=preposition; PERF=perfective; PTV=perfect; POS= possessive; PRES=present; PP=personal pronoun; RVRS =reversive; SM=subject marker; SEP=separative.

[2] Unless otherwise indicated, Kuria, Ngoreme, Simbiti and Gusii data are the result of data collection conducted by the author. Simbiti and Ngoreme data were gathered in Musoma (Tanzania) October–December 2015. Gusii and Kuria data were gathered in Rongo and Kisii (Kenya) during the same time

(1) *I-soloondu* *kw-ambʊk-a* *rí-**rɪ*** *ʊ-kuta*
5-lizard INF-climb-FV SM5-AUX 14-wall
'The lizard will climb the wall.' (Rangi, Gibson 2012:110)

(2) *N-ko-nyaháár-ék-á* *á-**re***
FOC-INF-harm-STAT-FV SM1-AUX
'He is becoming injured.' (Kuria, Cammenga 2004:259)

This paper examines this unusual word order, with a view to understanding its possible origins. Here it is proposed that the verb-auxiliary order has its origins in a verb-fronting construction which was used historically to convey predicate focus. Evidence for this pathway of grammatical change comes from the synchronic distribution of these constructions within the languages, as well as from comparative cross-Bantu evidence which shows variant word order employed for pragmatic purposes. The account forwarded here adds support to the proposal by Hyman and Waters (1984) and Güldemann (2003) that the progressive aspect is an inherently focal category in Bantu and that there is a cline leading from focus marking to the encoding of certain tense-aspect combinations in the language family. It also shows the interplay between information structure and the encoding of tense-aspect-mood distinctions.

The paper is structured as follows: Section 2 introduces the data, exemplifying the verb-auxiliary order in the languages under investigation. Section 3 outlines the proposed pathway of grammatical change that gave rise to this word order, drawing on wider patterns of language change and grammaticalisation. Section 4 provides a note on the possible role of language contact in the development of this structure. Section 5 constitutes a summary and conclusion.

2 Verb-auxiliary order in East African Bantu

2.1 Auxiliaries in Bantu languages

The Bantu language family is made up of some 350-500 languages spoken across much of Central, Eastern and Southern Africa (Marten *to appear*; van de Velde and Bostoen 2019). Bantu languages exhibit broad typological similarities: dominant SVO word order, the presence of noun classes, and extensive systems of agreement the effects of which can be seen across both nominal and verbal domains. Bantu languages also exhibit highly agglutinative, primarily head-marking morphology and a rich verbal complex in which inflectional and derivational affixes join to an obligatory verb stem to convey subject and object information, as well as in some instances, polarity, deixis and focus. While not all elements are necessarily present in a given verb form, those that do appear must do so in a rigid order.

Bantu languages employ a combination of simple and compound auxiliary verb forms to encode tense-aspect-mood distinctions (Nurse 2008; Nurse and Philippson 2006). Simple verb forms comprise of a single verb which may be inflected for tense and/or aspect information, as can be seen in the example from the East African Bantu language Swahili in (3).

(3) *Ni-na-end-a* *shule-ni*
SM1SG-PRES.PROG-go-FV 9.school-LOC
'I am going to school.' (Swahili)

Compound constructions involve at least one auxiliary form and a main verb. For the purposes

period. Rangi data are taken from Gibson (2012). Swahili examples are based on the author's own knowledge. In some instances, glosses have been adapted to aid presentation.

of the current discussion, auxiliaries are considered to be elements which are verbal in origin, but which make a functional rather than a lexical contribution to the clause.[3] Auxiliaries in the Bantu languages are commonly associated with functional categories such as tense, aspect and mood, and negation (Gibson 2019).

A range of grammaticalisation processes have resulted in variation in Bantu auxiliary constructions. Languages differ with regard to subject-marking properties, the distribution of tense-aspect-mood information across the elements of the clause and the number of auxiliaries present in the language. In some languages, both the auxiliary and the main verb exhibit subject agreement ((4)-(5)), while in other languages only the auxiliary hosts subject information and the main verb appears in an uninflected or infinitival form (6).

(4) *Wa-li-kuwa* *wa-me-fik-a*
SM2-PAST-AUX SM2-PERF-arrive-FV
'They had arrived.' (Swahili)

(5) *Re-tlo-b-e* *re-rek-ile*
SM1PL-FUT-AUX-SBJV SM1PL-buy-PERF
'We will have bought' (Northern Sotho, Nurse 2008:157)

(6) *Ti-na* *ku-gúl-a*
SM1PL-AUX.PAST INF-buy-FV
'We were buying.' (Chichewa, Nurse 2003:91)

Despite the micro-variation found in auxiliary constructions across the language family, pre-verbal auxiliary placement predominates across the Bantu languages (Nurse 2008, Gibson 2019). The six languages under examination in the current study – Rangi, Mbugwe, Gusii, Kuria, Simbiti and Ngoreme – represent an exception to this pattern and all exhibit verb-auxiliary order as a regular part of their tense-aspect systems.

The presence of the verb-auxiliary order in these languages gives rise to the following questions: Assuming that these structures have their origins in the more common auxiliary-verb order, how did they come into being? What process(es) of change gave rise to this divergent word order against the backdrop of an otherwise dominant auxiliary-verb order across Bantu?

The next sub-section presents data exemplifying this construction in the six languages, before the historical development of the construction is addressed in Section 3.

2.2 Verb-auxiliary order: The data

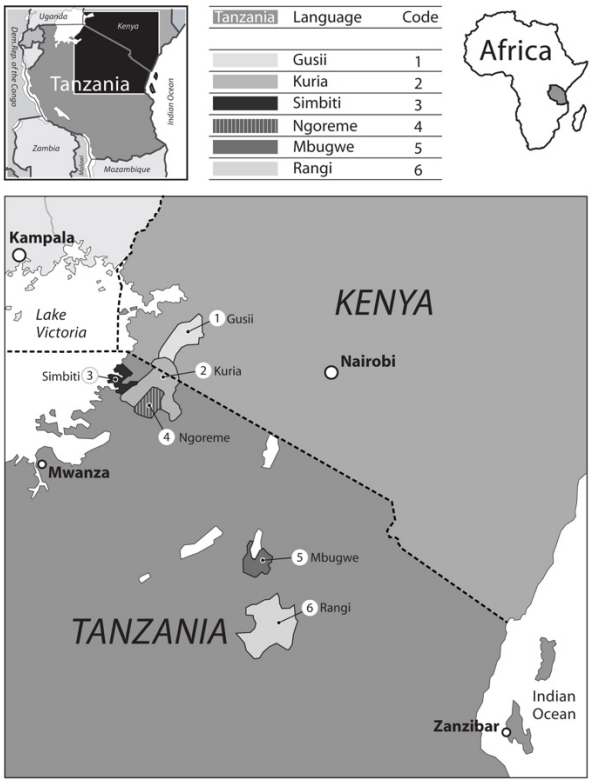
The languages in which verb-auxiliary order has been identified are all spoken in East Africa.[4] As can be seen on examination of the map below, the languages fall into two groups:

[3] For the purposes of the current study auxiliary forms are those that make no independent lexical contribution to the clause in which they appear (and which as a result can only appear with a lexical main verb). However, since in Bantu languages main verbs are common sources for auxiliary forms, there are also 'auxiliarising' forms which occupy an intermediate stage in this process or appear synchronically as both auxiliary forms and main verbs in a single language. An example of this can be in Swahili where the verbs *-ja* 'come' and *-pata* 'get' have restricted auxiliary usage and can only occur with infinitival complements (cf. Ashton 1947: 273–277).

[4] As pointed out by an anonymous reviewer, there are also varieties of Kikongo which appear to allow post-verbal placement of the auxiliary. These constructions – termed the locative-infinitive construction – are not treated under the current examination due to their locative component. However, it is worth noting that some of the broader observations relating to progressive aspect and

Rangi and Mbugwe which are spoken in central Tanzania and Gusii, Kuria, Simbiti and Ngoreme which are spoken near Lake Victoria, covering areas in both Kenya and Tanzania.

Although they are all Bantu languages, they do not all form a coherent sub-set of closely related languages, and not all six are spoken in geographic proximity to one and other. Rangi and Mbugwe are closely related and although they are no longer in direct contact with each other, are assumed to have migrated to this area of central Tanzania together (Kießling et al. 2008). Gusii, Kuria, Simbiti and Ngoreme are spoken in a contiguous area. However, Kuria is the only language which is in direct contact with all of the other three languages of this area. Crucially, they are all in contact with other languages spoken in the area, including Bantu and non-Bantu languages (see Section 4 for additional details of the linguistic ecologies in which these languages are found).



Despite the differing degrees of relatedness and geographic proximity, these languages are united by the presence of the verb-auxiliary order as a regular part of their tense-aspect systems. The data showing the verb-auxiliary constructions in each of the languages is presented in turn below.

predication focus which will be explored in the current paper also apply in these Kikongo constructions (see de Kind et al. 2015 for more on this).

2.3 Rangi

Rangi ([lag], F33) is spoken in the Kondoa region of central Tanzania.[5] Verb-auxiliary order is found in Rangi in two tenses – the immediate future tense and the general future tense. Both of these tenses are formed using an infinitival verb and an inflected auxiliary. While in many Bantu languages, infinitival verb forms obligatorily appear with the class 15 prefix, in Rangi bare, prefixless infinitival verb forms are also widely attested. In the absence of the class 15 prefix therefore, the Rangi infinitive can be identified on the basis that it does not carry any tense-aspect-mood information. The immediate future tense is typically used for an action or event which is imminent or will occur in the very near future, and is formed using the auxiliary *-íise*. The infinitive can appear either with the infinitival prefix *kʊ-* (or *kw-* before a vowel) (7) or in a 'bare' form without this prefix (8).[6] Crucially, in these contexts, an attempt at pre-verbal auxiliary placement results in ungrammaticality (9).

(7) *Kw-i-súm-ʊl-a* *n-íise* *i-hi* *mbʊ́ri* *haaha.*
INF-OM9-take-SEP-FV SM1sg-AUX 9-DEM 9.goat now
'I will take this goat now.'

(8) *Wul-a* *n-íise* *ma-taanga* *sa* *weéwe* *dee* *joo-terek-a.*
buy-FV SM1SG-AUX 6-pumpkins for 2SG.PP then DIR-cook-FV
'I will buy a pumpkin for you to come and cook.'

(9) **N-íise* *térek-a* *chá-kʊrya*.
SM1SG-AUX cook-FV 7-food
Intd. 'I will cook food.'

The general future tense is used to encode an action or event that will occur at an unspecified point in the future or in the more distant future. The general future tense is formed using an infinitival verb and the auxiliary *-rɪ*, as can be seen in examples (10) and (11) below.

(10) *Mama* *jót-a* *á-rɪ* *maaji* *mpolɪ.*
1.mother get.water-FV SM1-AUX 6.water later
'Mother will get water later.'

(11) *Ki-lwire* *ɪ-kɪ* *kwa-n-jul-a* *ki-rɪ.*
7-illness DEM-7 INF-OM1sg-kill-FV SM7-AUX
'This illness will kill me.'

[5] Following Guthrie (1967-71) and the update provided in Maho (1999), Bantu languages are commonly referred to using a system of zones which are assigned a letter. The languages in each zone are further distinguished by a number. Thus, Rangi – F33 – belongs to zone F and is geographically proximate to F34 and so on. This system is based on geographic rather than genetic classification.

[6] The factors determining the presence of absence of the class 15 prefix on the Rangi infinitive remain unclear and further research is required to fully understand its distribution. Stegen (2006) proposes that the presence of the *kʊ-* prefix encodes a certain future whilst is absence is associated with the 'uncertain future'. Gibson (2012) concluded that the presence or absence of the prefix is also determined by a number of other factors, including whether the verb stem is vowel-initial or monosyllabic, the nominal or verbal status of the infinitive. However, Rangi is certainly not alone amongst Bantu languages in exhibiting both a prefixed and a prefixless infinitival form, with similar alternations also observed in Sambaa (Riedel 2009:26), Lusaamia (Botne 2004) and varieties of Kikongo (de Kind et al. 2015).

It is worth noting however that whilst these two future tense constructions exhibit verb-auxiliary order, all other tenses in Rangi which are formed through compound verbal constructions exhibit the more Bantu-typical auxiliary-verb order, as exemplified by (12) and (13) below.

(12) *V-íja* *va-tɪɪte* *ɔ-jusi* *w-óó* *rut-a*
SM2-AUX.PAST2 SM2.PERF-have.PERF 14-knowledge 14-of forge-FV
'They used to be blacksmiths'
lit.: 'They used to have the knowledge of forging.'

(13) *N-áá-rɪ* *n-a-tey-ire* *mɔ-teho* *w-aanɪ*
SM1SG-PAST1-AUX SM1SG-PAST1-set-PERF 3-trap 3-my
'I have set my trap.'

2.4 Mbugwe

Mbugwe ([mgz], F34) is spoken in the Babati District of northern central Tanzania. Babati – the centre of the Mbugwe-speaking community, is some 60km away from Kondoa Town, the heart of the Rangi-speaking community. Rangi and Mbugwe are closely related and are thought to originate from a common predecessor language. In Mbugwe, tense-aspect distinctions are encoded through a combination of markers which appear in the pre-stem position and inflectional endings which appear after the verb stem, as well as through the associated tone marking (Wilhelmsen 2014). A past tense event, for example, is conveyed using a simple verbal construction in which the past tense prefix *áa-* and the suffix *-á* are joined to a verb stem (14), whilst a future tense construction is formed using the prefix *já-* (15).

(14) *N-áa-rem-á* *yonda* *r-áne* *wiki* *aloká.*
SM1SG-PAST-cultivate-PAST 5.farm 5-my week past
'I cultivated my farm last week.' (Mous 2004: 8)

(15) *Si-já-lósék-a* *na* *wée.*
NEG.1SG-FUT-talk-FV PREP you
'I will not talk with you.' (Mous 2004: 9)

However, compound constructions in Mbugwe regularly exhibit verb-auxiliary order. This ordering is found in Mbugwe in six different tense-aspect combinations and a different auxiliary is associated with each tense-aspect combination. These constructions are all formed using an inflected auxiliary form and a main verb. The recent past progressive, for example, is comprised of an infinitival verb form alongside the auxiliary *-re* and the past tense prefix *á-* (16).

(16) *O-rem-a* *n-á-re* *i-onda* *re-ááné*
INF-cultivate-FV SM1SG-PST-AUX 5-field 5-1SG.POSS
áfá *á-a-fik-á.*
16PP.DEM.PROX SM1-PST-arrive-P3
'I was cultivating my farm when he arrived.' (Wilhelmsen 2014: 3)

In Mbugwe, different auxiliary forms are used in different tenses. Thus, the present imperfective is formed using the auxiliary *-kéénd*ɛ́ (17), the habitual is formed using the auxiliary *-áándá* (18) and the future perfective is formed using *-je* (19).[7]

(17) *I-onda* *o-rem-w-a* *re-kéénd*ɛ́ *ne* *mo-ntomoká*.
5-field INF-cultivate-PASS-FV SM5-AUX.PRS CONN 1-woman
'The farm is being cultivated by the woman.' (Wilhelmsen 2018: 124)

(18) *Na* *o-mo-túmb-a* *n-áándá* *i-jóva* *baa* *éensiko*.
CONN INF-OM1-follow-FV SM1SG-AUX.HAB1 5-god even today
'I am following God even until today.' (Wilhelmsen 2018: 130)

(19) *O-sír-a* *ko-je* *na* *va-ána*.
INF-finish-FV SM1PL-AUX.FUT CONN 2-child
'We are going to die, and the children too.' (Wilhelmsen 2018: 145)

The past progressive employs the auxiliary *-áyse* (20), while the habitual employs the auxiliary *-jéénde* (21).

(20) *Na* *o-kól-a* *n-áyse* *ny-ááfu* *nɛ́ɛ*
CONN INF-take.up-FV SM1SG-AUX.PROG2 10-NET 1SG.PP
n-túr-iryɛ o-kól-a *ny-ááfu* *i-veeré*
'I was taking up nets, and I had already taken up two nets…' (Wilhelmsen 2018 :129)

(21) *Hamu* *ne* *ó-tɔ́ɔ́l-a* *o-jéénde* *nkúúnda* *monɔ́* *o-tiingok-a*
Look FOC INF-get-FV SM2SG-AUX.HAB 9.problem alot INF-walk-FV
na *molo* *óko*
CONN 6-foot 17.DEM.PROX
'Because you have big problems walking here on foot.' (Wilhelmsen 2018 :130)

In contrast to Rangi, in which there are compound constructions in different tense-aspect combinations which exhibit the more Bantu-typical auxiliary-verb order, all auxiliary constructions in Mbugwe exhibit verb-auxiliary order.

2.5 Gusii

Gusii ([guz], E42) is spoken in Kisii District of Nyanza Province in western Kenya. Gusii borders the Kuria-speaking area in Tanzania, as well as being in contact with a number of non-Bantu languages in Kenya. In Gusii, the verb-auxiliary order is found in four tenses: present continuous, habitual, hodiernal past continuous and recent past continuous. All four of these constructions are formed using the auxiliary *-ré*, with the addition of a different tense-aspect marker in each instance.

The present continuous is formed using the infinitive and the auxiliary *-ré*. The infinitive carries the prefix *kó-*, while the auxiliary conveys the subject information. The infinitive also hosts the prefix *ne-* which has been described for Gusii (and across a range of Bantu languages,

[7] In Mbugwe, the different auxiliary forms appear to be more transparently linked to the verbs in which they are presumed to have their origins. The future tense auxiliary *-je* for example being derived from the verb *-ja* 'come'.

see §3.3.2) as a focus marker (Cammenga 2002). As can be seen in the examples below, the auxiliary appears post-verbally in these constructions (22)-(24). [8]

(22) *Maria ń-kó-ges-a á-ré o-bo-ori.*
Maria FOC-INF-harvest-FV SM1-AUX AUG-14-millet
'Maria is harvesting finger millet.'

(23) *Ń-kó-gend-a ńdé e-chiiro.*
FOC-INF-go-FV SM1SG.AUX AUG-9.market
'I am going to market.'

(24) *Ń-kó-riing-or-a ńdé a-ma-raangeti.*
FOC-INF-fold-RVRS-FV SM1SG.AUX AUG-6-blankets
'I am unfolding the blankets.' (Cammenga 2002: 385)

The habitual construction is also formed using the auxiliary *-ré* and a verb form which is prefixed with the infinitival marker *kó-* and the focus marker *n-*. However, in addition to these elements, the occasional habit hosts the marker *-ráá* (25). Cammenga (2002) does not assign a specific meaning to the tense-aspect maker *-ráá*, however on the basis of the translation provided it appears to convey continuous aspect.

(25) *Ń-kó-ráá-rut-á ńdé.*
FOC-INF-TAM-bite-FV SM1SG.AUX
'I bite (occasionally).' (Cammenga 2002:488)

The hodiernal past continuous is also formed using an infinitival verb and the auxiliary *-ré*. However, in this tense the auxiliary hosts the hodiernal past tense prefix *áá-*, as illustrated ine example (26).

(26) *Ń-kó-rut-a mw-áá-ré.*
FOC-INF-bite-FV SM2PL-PAST1-AUX
'You were biting (earlier today).' (Cammenga 2002:493)

In the recent past tense, the auxiliary *-ré* hosts the recent past prefix *áa-* and the habitual suffix *-nge*, as can be seen in example (27) below.

(27) *Ń-kó-ges-á n-áa-renge*
FOC-INF-harvest-PAST SM1SG-PAST2-AUX.HAB
'I was harvesting.' (Whiteley 1959:34)

2.6 Kuria

Kuria ([kuj], E40) is spoken in the Mara region of northwest Tanzania and Nyanza Province in southwest Kenya. The Kuria-speaking communities border all three of the other verb-auxiliary languages that are spoken in the Lake Victoria area, i.e. Simbiti, Ngoreme and Gusii (see Map 1).

[8] In these examples from Gusii, the combination of the first person singular subject marker *n-* and the auxiliary *-re* is realised as *nde* as a result of a process of fortition.

In Kuria, verb-auxiliary order appears only in the present progressive tense.[9] The construction is formed through the use of an infinitival verb form and the auxiliary *-re*. The infinitive hosts the class 15 marker *ko-* and the focus marker *n-*, while the auxiliary is inflected for subject information, as in examples (28)–(30) below.[10]

(28) *N-ko-nyaháár-ék-á á-re*
FOC-INF-harm-STAT-FV SM1-AUX
'He is becoming injured.' (Cammenga 2004:259)

(29) *N-ko-rim-a bá-re ba-nyor-e i-bi-yakoria*
FOC-INF-farm-FV SM2-AUX SM2-get-SBJV AUG-8-food
'They are farming in order to get food.'

(30) *Hano tu-hik-ire o-mo-gondo n-ku-busur-a tó-re*
when SM1pl-arrive-PERF AUG-3-farm FOC-INF-plant-FV SM1pl-AUX
a-ma-hindi
AUG-6-maize
'When we arrive at the farm we plant maize.'

In Kuria, it is also possible for the infinitival verb to appear without the focus marker *n-*. In this case, the infinitival prefix also hosts the augment vowel *o-*, resulting in the prefix *oku-*, as can be seen in example (31).

(31) *Taata oku-riisiy-a á-re i-chi-tuko*
Father INF-graze-FV SM1-AUX AUG-10-livestock
'Father is herding livestock.'

The existence of examples such as that shown in (31) in Kuria give rise to a number of questions related to the interpretation of these constructions. If the *n-* prefix is a focus marker, what is the difference between those forms that appear with the focus marker and those which appear without it? Indeed, this question can be extended to all four of the languages which host a focus marker – Gusii, Kuria, Ngoreme and Simbiti. This is an issue which will be returned to in further detail in Section 4.

2.7 Ngoreme

Ngoreme ([ngq] E40) is spoken in the Mara Region of north-west Tanzania. Ngoreme and Kuria are neighbouring languages (see Map 1). In Ngoreme, the verb-auxiliary order is found in the present progressive and the past progressive. Both of these constructions are formed using an infinitive and an auxiliary. In the present progressive the main verb hosts the infinitival prefix *ko-* and the focus marker *n-*, whilst the auxiliary *-ni* is inflected for subject information ((32)-(34)).

(32) *Reero n-ko-gi too-ni Tarisaramu.*
9.today FOC-INF-go SM1pl-AUX Dar-es-Salaam

[9] In Cammenga (2002) this construction is also termed the 'simple present tense'. Since there does not seem to be any morphological or semantic distinction between these two forms, for the purposes of the current work they are considered as a single form and in the current discussion termed the present progressive.

'Today we are going to Dar-es-Salaam.'

(33) *N-ku-ba-abek-er-a* *too-ní*
FOC-INF-OM2-build-APPL-FV SM1pl-AUX
'We are building for them.' (Lotta Aunio p.c.)

(34) *N-gu-tem-a* *naa-ní* *u-ru-siri* *ne-ke-gisu*
FOC-INF-cut-FV SM1SG-AUX AUG-11-rope COP-7-knife
'S/he is cutting a rope with a knife.'

The past progressive is formed through a similar compound construction to that employed in the present progressive. However, the auxiliary used in the past progressive is *-re*.

(35) *N-go-tuk-a* *tw-áá-re* *ri-roma bhono* *n-to-tig-ire.*
FOC-INF-dig-FV SM1PL-PST-AUX 5-hole now FOC-SM1PL-stop-PERF
'We were digging a hole, now we have stopped.' (Tim Roth, p.c.)

The verb-auxiliary order in Ngoreme is not associated with all auxiliary constructions but is tense-specific (the same was also seen to be true of Rangi, see Section 2.3). As such, there are also auxiliary constructions which exhibit the more typologically common pre-verbal auxiliary placement, as can be seen in example (36) below.

(36) *ba-a-re* *ba-tuk-ire*
SM2-PST-AUX SM2-dig-ANT
'They had dug.'

2.8 Simbiti

Simbiti ([ssc] E403) is spoken in the Tarime District of the Mara region of north-west Tanzania. The Simbiti-speaking community is also in direct contact with speakers of Kuria. Simbiti uses a combination of simple and compound constructions to express a range of tense-aspect-mood distinctions. In a number of these compound constructions, the auxiliary appears in the pre-verbal position. This can be seen in examples (37) and (38) below where the auxiliary *-re* appears before the verb in the past anterior form.

(37) *N-aa-re* *n-rooshe* *nyinkyo* *a-ha-sɛ.*
1SG-PST-AUX SM1SG-see.ANT 9.morning AUG16-16-place
'I saw a place in the morning.' (John Walker p.c.)

(38) *A-aa-re* *a-ishooyi* *ri-tonga* *i-mwe.*
SM1-PST-AUX SM1-be.full.CAUS.ANT 5-basket 5-one
'…he had filled one basket.' (John Walker p.c.)

However, verb-auxiliary order in Simbiti is found in the present progressive, present persistive and the present habitual tenses. All of these tense-aspect combinations are formed using a main verb followed by the auxiliary *-re* which carries subject information. In the present progressive, the verb appears with the infinitival prefix *ku-* and the focus marker *n-* ((39)-(40)).[11]

[11] In Simbiti, as was also seen in Gusii in §2.5 above, the combination of the first person singular subject marker *n-* and the auxiliary form *-re* is realised as *ndi* as a result of a process of fortition.

(39) *N-ku-tug-a ndi sɛ-ng'oko.*
FOC-INF-keep-FV SM1sg.AUX 10-chickens
'I keep chickens.'

(40) *Tw-a-hik-ire n-twa-nyoore ghooko n-kɔɔ-kɔr-a*
SM1pl-arrive-PST-arrive-ANT FOC-SM1PL-find.ANT1.grandma FOC-INF-do-FV
a-re a-ma-kara w-akɛ.
SM1-AUX AUG4-4-charcoal 1-his/her
'When we arrived, we found grandma making her charcoal.' (John Walker p.c.)

In the present habitual form, the verb appears with the habitual suffix *-anga* in addition to the prefix *ku-* and the focus marker *n-*. The auxiliary carries subject information and appears after the main verb in declarative main clauses ((41)-(42)).

(41) *N-ku-bhin-anga to-re.*
FOC-INF-dance-HAB SM1pl-AUX
'We are (usually) dancing.' (John Walker p.c.)

(42) *N-ku-ri-anga ndi u-bhu-kima na a-ma-harage.*
FOC-INF-eat-HAB SM1SG.AUX AUG-14-ugali CONN AUG6-6-beans
'I usually eat *ugali* and beans.'

The persistive is constructed using an auxiliary-based construction involving the combination of an inflected verb and the auxiliary *-re* which hosts the prefix *kee-*, as in (43).

(43) *N-ku-shumaash-a to-kee-re na-wɛ.*
FOC-PROG-speak-FV SM1PL-PER-AUX CONN-3SG.PP
'We still speak with him/her' (Walker 2013: 108)

Finally, although Simbiti remains a larger under-described language, on the basis of the available data it appears that there is also some dialectal difference in the presence of this construction. The verb-auxiliary examples provided above have been observed in the Simbiti-Kiroba-Iregi group of varieties which in addition to the compound verbal construction, is able to express the present tense thorugh a simple verb form, as can be seen in example (44).

(44) *Ba-ra-tuka*
SM2-CONT-dig-FV
'They are digging'

However, in the other main variety – the Hacha-Sweta group – this compound verbal construction is not found and only the simple verb form is used to express the present tense, as can be seen in example (45).

(45) *N-gu-tuka*
FOC-INF-dig-FV
'They are digging'

2.9 Summary

To summarise, the subset of six East African Bantu languages under examination here all exhibit verb-auxiliary order as a regular part of their tense-aspect-mood systems. Crucially, in these languages, the inverse order is unacceptable in these constructions (at least in declarative main clauses, a point that will be examined in §3.4.1) The tense-aspect combinations and the number of constructions in which this word order is found differ from language to language. An overview of these contexts is presented in Table 1 below.

Table 1: Verb-auxiliary constructions across the languages

	Rangi	**Mbugwe**	**Gusii**	**Kuria**	**Ngoreme**	**Simbiti**
No. of tenses	2	6	4	1	2	3
Tense-aspect	Immediate future General future	Present imperfective, Habitual Past progressive, Future perfective	Habitual, Present continuous, Past continuous, Past habitual	Present progressive	Present progressive, Present habitual	Present progressive, Past progressive
No. of auxiliaries used	2	6	1	1	2	1
Form of auxiliaries	*-rɪ* *-íise*	*-re* *-je* *-áyse* *-áandá* *-jéénde* *-kééndé*	*-re*	*-re*	*-re* *-ni*	*-re*

As can be seen upon examination of the table above, the languages vary with regard to the number of tenses in which the verb-auxiliary order is found, as well as the number of auxiliary forms used in the formation of these constructions.

Mbugwe is the language in which the verb-auxiliary order is found in the most tense-aspect combinations – six. Kuria lies at the other end of the spectrum, with the verb-auxiliary order found in just one tense-aspect. In some of the languages, all auxiliary constructions exhibit this order (e.g. Mbugwe), whilst in other languages both the verb-auxiliary order and auxiliary-verb order is found although the different orders are associated with different tenses (e.g. Rangi). Despite this variation, a number of generalisations can still be seen to hold across the subset. A (variant) of the copula form *-re* is involved in the formation of at least one of the constructions in all six languages. The copula *-re* can be traced back to the Proto-Bantu reconstructed copula **-dì* (Guthrie 1967-71 Vol. III: 150, C.S. 547; BLR3 940) or **-di* (Meeussen 1967: 86). Ngoreme additionally employs the auxiliary *-ni*. In Mbugwe, the six different auxiliary forms can be seen to be more transparently derived from lexical verbs (for example the auxiliary *-je* appears to be derived from the verb *-ja* 'come').

A further generalisation is that all of the constructions involve either progressive or habitual aspect, or future tense interpretation. This is an observation which will be examined in further detail in Section 3.2 which explores the role of the encoding of tense and aspect in the process of grammatical change that is proposed to have given rise to this word order.

3 The development of the verb-auxiliary order

3.1 Background

The assumption made here is that the verb-auxiliary order found in these languages has its historical origins in auxiliary-verb order. This is based on the observed prevalence of auxiliary-verb order across East African Bantu (Nurse 2003; Gibson 2019, Marten et al. 2018) and in SVO languages (Greenberg 1966; Dryer 1992).[12] This proposal is also supported by synchronic data from these six languages – all of which also exhibit auxiliary-verb order in certain tense-aspect combinations and/or syntactic contexts (see §3.1). The question is therefore, what historical process of change may have given rise to the verb-auxiliary order found in these languages?

At least for Rangi and Mbugwe, it has been proposed that the verb-auxiliary order is the result of contact with non-Bantu languages (Stegen 2002, Nurse 2003, Dunham 2005, Gibson 2013). This is a point that will be returned to in Section 4. However, this paper represents the first time that a systematic study of this word order has been conducted on the basis of data from six languages, and the account developed here is that this word order is primarily the result of language-internal processes of grammaticalization, albeit in a context of sustained contact with non-Bantu languages.

Here it is proposed that the presence of the verb-auxiliary order found in these languages is related, at least historically, to information structure. Specifically, that the verb-auxiliary order started out as a way of encoding predication focus. This structure subsequently grammaticalized to encode progressivity – an aspect which has been considered to be 'inherently focal' in Bantu (Hyman & Watters 1984; Güldemann 2003). However, before outlining the pathway of change in further detail, it is necessary to establish the terminological background on which the subsequent proposal is based.

The notion of 'predication focus', as defined by Güldemann (2003: 330–331) is used to refer to focus centred on the predicate. This crucially excludes objects and adjuncts and is therefore distinct from Lambrecht's (1993: 226) broader notion of predicate focus which targets the entire verb phrase, including objects and adjuncts when present. In contrast, therefore, Güldemann's (2003) conceptualisation of 'predication focus' is narrow in that it is centred on the verb lexeme or on a predication-operator associated with the verb expressing polarity or truth for example, or as will be seen in the subsequent discussion, tense or aspect. This also contrasts with the notion of verb focus, in which focus is on the meaning of the verb itself.

3.2 The proposed pathway of change

The proposal is that historically, while auxiliary-verb order was the dominant structure, verb fronting was possible in order to convey predication focus (as is still the case in a number of related languages, see §3.3.). This verb-auxiliary order subsequently developed to become the standard way to encode certain tense-aspect combinations, as is attested today. The proposed pathway of change is outlined in (46) and described below.

(46) **Stage I** **Stage II** **Stage III**

[12] As was pointed out by an anonymous reviewer, it is not always the case that the Verb-Auxiliary order in SVO languages is historically derived from Aux-Verb order. It is possible for a language that exhibits SVO order to be the diachronic result of a SOV order (as was the case in the development of Romance languages from Latin, for example), in which the inflected Aux tends to follow the Verb (see Greenberg's linguistic universal 16). Whilst this is an important distinction cross-linguistically, for the Bantu languages it seems reasonable to assume that the historical order was indeed Auxiliary-Verb.

Auxiliary-verb > Auxiliary-verb ~ (*ni* cleft) verb-auxiliary > verb-auxiliary

In Stage I, auxiliary-verb order is the standard order found across all auxiliary constructions in these languages. Stage II involves the option of the verb appearing before the auxiliary for pragmatic purposes. This verb-auxiliary order would have been used to convey predication focus, emphasising either the verb lexeme or values such polarity, or TAM distinctions. This means that at some point in the historical development of the languages, both auxiliary-verb and verb-auxiliary order would have possible, albeit with different information structural properties associated with the two orders. At this stage, the verb-auxiliary structure is reanalysed as a marker of progressive aspect. This stage may have been facilitated by the close relationship between progressive marking and focus, particularly the inherently focal status of progressives in Bantu.

In Stage III, the process of grammaticalization involves a move from a verb focus construction to a focused progressive and ultimately to a simple progressive which is unmarked in information structural terms. Since the predication focus reading was lost, the construction is also open to other information structures that may be superimposed. At this point, the construction in these six languages ceases to be complex and is reinterpreted as a single clause.[13] This also fits with the cross-linguistic observation that biclausal structures often develop into monoclausal structures as part of processes of grammaticalisation, as is seen with clefts structures and auxiliary constructions more broadly (Harris & Campbell 1993:152, 172)

A crucial part of this explanation is the surface ambiguity historically associated with the verb-auxiliary order (at Stage II), under which it could be analysed as either a means for conveying predicate focus or a progressive aspect. This links to the more broadly observed polysemy of markers and strategies for expressing progressive aspect and predication focus across Bantu (Güldemann 2003; Hyman & Waters 1984). In Rangi, this also sees the development of the focused progressive formed with the auxiliary 'go' to a simple immediate future and either at the same time or by analogy with the immediate future, the development of the general future tense (see §3.5).

This change from Stage II to Stage III would have represented a process of reanalysis, under which the underlying structure of the syntactic contribution made by the verb-auxiliary order would have changed, albeit without any change in morphological realisation (Harris & Campbell 1995). Crucially this would have involved a change in the grammatical function of the verb-auxiliary order, with the word order initially used to convey predication focus (i.e. a pragmatic function) to it encoding a progressive aspect (i.e. a semantic function). This means that the secondary progressive meaning subsequently became more strongly semanticised than its source meaning as an indicator of predication focus.

Two interlinked grammaticalization paths can therefore be observed. The first involves the grammaticalization of a construction that conveys predication focus. The second involves the levelling of the information structure of this construction over time, and the conventionalisation of the structure previously associated with the predication focus – i.e. verb-auxiliary order – which ultimately ceased to encode predication focus and merely conveyed the content of the proposition. [14]

[13] This differs from the Kikongo case where elements can intervene between the verb and the auxiliary.

[14] A similar origin for progressive aspect has also been noted in other unrelated languages. In Yucatec Maya, for example, the structure that had previously been associated with verb focus has been maintained in the language but is no longer associated with a focal reading (Lehmann 2008).

This pathway of change, is supported by observations from four domains: i) Comparative insights from Bantu languages in which verb doubling is attested; ii) the distribution of the prefix *n-* and its relation to cleft construction; iii) the synchronic alternation between pre- and post-verbal auxiliary placement; iv) the specific tense-aspect distinctions with which these constructions are associated. These characteristics can be examined in terms of support for the proposed pathway based on form on the one hand and meaning on the other and are discussed in turn below.

3.3 Support for the pathway of change: Form

3.3.1 Link between verb-auxiliary order and verb fronting

The first insight into the pathway of change proposed here comes from comparative observations from Bantu languages which exhibit verb fronting constructions. The use of verb fronting is widespread throughout Bantu and was observed by Meeussen (1967:121) who reconstructed verb fronting for Proto-Bantu and described it as the 'advance verb construction'. Such constructions typically involve a fronted infinitive or non-finite form and an inflected verb form. This can be seen in the examples from Kaamba below, where example (47) shows the simple verb form and (48) shows a verb doubling construction.[15]

(47) *wà-mú:-sàlá*
SM1-PROG-work-FV
'*il est en train de travailler.*'
'He is working.' (Kaamba, Bouka 1989 cited in Hadermanm 1966:160)

(48) *Sàl-á* *kà-mú-:sàl-á*
INF.work-FV SM1-PROG-work-FV
'*il est en train de travailler.' 'Travailler il est en train de travailler'*
'He is working.' Lit.: 'Working he is working.' (Kaamba, Bouka 1989 cited in Hadermanm 1966:160)

While both the examples in (47) and (48) express a present continuous event, the description for the verb fronting construction in (48) is that it is used to 'reinforce the concept of repetition in the carrying out of the action' (Hadermann 1996:160).[16]

A similar phenomenon has been described in number of varieties of Kikongo where the verb fronting construction is used to express focus on the lexical meaning of the verb, the TAM marker, truth value or what de Kind et al. (2015) more broadly describe as 'event-centrality'. As can be seen in the example in (49) below, the use of the verb fronting construction conveys a focal reading on the verb lexeme *sonika* 'write'.

(49) *Sónik-a kaka ba-sonik-idi*
write-FV only SM2-write-PRF
'They only WROTE a report' (Kimbeko, de Kind et al. 2015:115)

A similar situation can be seen in Suundi (H13b) where there is a distinction in the future tense between an unmarked verb form (50) and the verb fronting construction which is used to convey predication focus (51).

[15] In Kaamba there is an alternation between the canonical class 1 prefix *u-* and its variant *kà -*
[16] Original: *renforcer l'idée de repetition dans le déroulement de l'action'* (Hadermann 1996:160).

(50) *ndy-èká-tá:ng-à*
SM1sg-FUT-read-FV
'*je vais lire.*'
'I will read.' (Suundi, Haderman 1996:162)

(51) *kù-tá:ng-à* *ndy-èká-tá:ng-à*
15-read-FV SM1sg-FUT-read-FV
'I am going to READ.' (Suundi, Hadermann 1996:162)

The verb fronting construction is widespread among West Bantu languages of Guthrie's zones B and H (Hadermann 1996, de Kind et al. 2015). However, it has also been reported in the South-Western Bantu language Thimbukushu (Güldemann 2003:336) and in the Eastern Bantu languages Fwe (Gunnink 2016) and Gikuyu (Morimoto 2013, 2016, 2017).

The verb-auxiliary construction discussed in this paper does not involve verb fronting, since the main verb only occurs once. However, the presence in the Bantu family of constructions in which an infinitive appears before an inflected verb form to convey predication focus constitutes important cross-Bantu support for the idea that the verb-auxiliary order has its origins in a variant word order that was employed for pragmatic purposes.

Moreover, none of the languages which exhibit the verb-auxiliary order also exhibit verb doubling constructions akin to those discussed in this subsection. Those languages which exhibit verb doubling do not allow for post-verbal placement of the auxiliary. Crucially, the verb doubling construction involves repetition of the same verb form, not a lexical verb and an auxiliary.[17] However, even with the verb fronting construction there is a link between progressive aspect and predication focus, where in Kamba, for example, verb fronting is associated with progressive aspect.

3.3.2 Evidence from ni

The second observation is that in four of the six verb-auxiliary languages, these structures also include some variant of the marker *ni.* The marker *ni* and its variants have been described as focus markers, both specifically in the languages under examination here (Aunio et al. 2019*,* Cammenga 2002, Cammenga 2004; Walker 2013:156) and in the broader literature on Bantu languages (Schwarz 2003, 2007; Abels & Muriungi 2008).

In Gikuyu (E51), the same element – *ne* – is used as an identificational copula and to convey focus. In Gikuyu, the post-verbal position is the unmarked focal position (52). However, *ne* can be placed before a verb form in order to convey predication focus (53) or before a nominal to convey focus on the noun (54). The form *ne* also functions as the identificational copula (55).

(52) *à-kàà-gwr̀à* *nhàmà*
SM1-FUT-buy meat
'He will buy MEAT.' (Gikuyu, Bennett 1985:168)

(53) ***Ne*** *gu-thodék-a* *a-ra-mé-thodék-ire.*
FOC INF-fix-FV SM1-PST-9-fix-PFV
'He FIXed it. (Gikuyu, Morimoto 2013:9)

(54) ***Ne*** *mae* *Abdul* *a-ra-nuy-irɛ.*

[17] This gives rise to a broader comparative question as to whether verb doubling and verb fronting constructions are in complementary distribution. However, a more comprehensive understanding of verb doubling across Bantu would be needed in order to make any conclusions in this regard.

FOC 6.water 1.Abdul SM1-TAM-drink-ASP-FV
'It is WATER Abdul drank.' (Gikuyu, Schwarz 2003:17)

(55) *Karioki* ***ne*** *mo-rutani.*
Karioki COP 1-teacher
'Karioki is a/the teacher' (Gikuyu, Schwarz 2003:19)

Consider also the case of Haya (JE22). In Haya, *ni* functions as an identificational copula (56). However, a comparison of examples (57) and (58) also shows that the presence of *ni-* on the verb form conveys the present progressive.

(56) ***Ní*** *káto*
COP Kato
'It is Kato.' (Haya, Hyman and Watters 1984:260, 261)

(57) *ba-mu-kóma*
SM2-OM1-tie
'They tie him up.' (Haya, Hyman and Watters 1984:260, 261)

(58) ***Ni****-ba-mu-kóma*
PROG-SM2-1OM-tie
'They are tying him up.' (Haya, Hyman and Watters 1984:260, 261)

Finally, in Kamba (E55), the same element is used to mark predicate focus and to mark progressive aspect. Again, the unmarked focus position is the post-verbal position, as can be seen in example (59). However, predicate focus can be conveyed through the use of the focus particle/prefix *nĩ-* which appears before the predicate (60). However, in a construction such as (61), the use of *nĩ-* conveys a progressive reading.

(59) *Tũ-tônya* *kŵikwa* *ma-ũndu* *ásu*
SM1pl-can.PRES INF.do 6-things those
'We can do THOSE THINGS (not others)' (Kamba, Ndumbu and Whiteley 1962:174, 176, 179)

(60) ***N'ĩ-t'tônyá*** *kŵika* *ma-ũndu* *ásu*
PF-SM1pl-can.PRES INF.do 6-things those
'We CAN do those things.'

(61) ***N'ĩ-méũ-theka***
PROG-SM2-laugh
'They are laughing (they are about to laugh)'

Comparing these three language types we can observe the following: 1) a language in which the same element is used as an identificational copula and for predicate focus (Gikuyu), 2) a language in which the same element is used to mark progressive aspect and as an identificational copula (Haya), and 3) a language in which the same element is used to mark predicate focus and progressive aspect (Kamba).

Güldemann (2003) proposes that the element *ni* in both its preverbal and its prenominal use is a reflex of the same morphosyntactic structure. The element *ni* can therefore be considered to encode the reading 'it is (the case that X)'. Under this account, the polysemic nature of the

marker *ni* is taken as evidence that this is the direct result of a historical change from one to the other. Furthermore, rather than being an accidental occurrence, since it has been proposed that the progressive is an inherently focal category (Hyman and Watters 1984, Güldemann 2003), the relationship between these two elements is considered to be part of a process of grammaticalisation which sees the development of a form marking predicate focus into a marker of present progressive.

The languages under examination in the study differ somewhat in terms of the status of *ni* as a focus marker. Firstly, in Rangi and Mbugwe, there is no remnant of the marker *n-* on the verb-auxiliary construction. However, it is noteworthy that in both of these languages the infinitival can appear without the common Bantu infinitive noun prefix *ku-*. The precise factors determining the presence versus absence of *ku-* remain to be identified. However, a process of prefix reduction which also effects the infinitive noun class prefix has been noted to be common across Bantu languages (Bostoen and de Schryver 2015), including in Kikongo (de Kind et al. 2015). It may well be that any potential trace of the marker *n-* does not appear in these languages for independently motivated reasons – i.e. since the entire prefix is 'lost' there is nothing to host the *n-* marker.

In Gusii, Kuria, Ngoreme and Simbiti, the verb-auxiliary structures all have the verbal prefix *n-*. This can be seen in the examples from Gusii (repeated from (24) above) and Kuria (repeated from (28)) below.

(62) *Ń-kó-riing-or-a ńdé a-ma-raangeti.*
FOC-INF-fold-RVRS-FV SM 1sg.AUX AUG-6-blankets
'I am unfolding the blankets.' (Gusii, Cammenga 2002:259)

(63) *N-ko-nyaháár-ék-á á-re*
FOC-INF-harm-STAT-FV SM1-AUX
'He is becoming injured.' (Kuria, Cammenga 2004:259)

For Kuria at least, it has been noted that what may have historically been a focus marker appears to have become lexicalised in certain tenses and has 'lost its emphatic force' (Cammenga 2002:385). For the other languages there is no explicit statement to this effect. However, there does not appear to be an equivalent 'unfocused' corresponding form without the marker *n-*. Nor do there seem to be specifically focal readings associated with these constructions from a synchronic perspective. This suggests that what may have begun life as a focus marker, is now a fully grammaticalised feature of these constructions that does not convey particular information structural effects.

The form *ni* (and its variants) is still found however, as a predicative copula and in the formation of cleft structures. This can be seen in Gusii for example where *ni* appears at the front of content interrogatives (formed using a cleft construction) (64). In Ngoreme, the copula *no* is used in non-verbal predication (65), as well as in the formation of a cleft construction and to convey term focus.

(64) *Ni-nki ki-aa-re ko-gwera enyoomba?*
COP-what 7-CONT-AUX INF-fall 9.house
'What is falling on the house?' (Gusii, *author fieldnotes*)

(65) *Eni no-mu-mwarimu*
1sg.PP COP-1-teacher
'I am a teacher.' (Ngoreme, *author fieldnotes*)

(66) *Ni ne n-gó-sóm-a*
COP 1sg.PP FOC-INF-read-FV
'It is ME who is reading' (Ngoreme, *author fieldnotes*)

(67) *Ono m-mɔrɛmi*
DEM.1 COP-1-farmer
'This is a farmer.' (Simbiti, John Walker p.c.)

Additional research would be needed to better understand the intricacies of the information structure associated with the copula *ni* in all of these languages from a synchronic perspective. However, here the presence of the marker *n-* on these forms is considered as evidence in support of the idea that the verb-auxiliary construction in these four languages went through a stage of transition (Stage II in the schema presented in §3.2) in which the verb-auxiliary order was part of a cleft construction. This cleft construction would have been used to covey either term focus (i.e. focus on a noun phrase) when the clefted element was a nominal or to convey predication focus (i.e. focus on the verb lexeme or the relevant predicate-operator). The close semantic and pragmatic link between predication focus and progressive aspect, may also have resulted in the use of these clefted constructions with progressive forms more often than with other tense-aspect combinations, thereby leading to the grammaticalization of this structure overall.

While we do not have access to historical records of the intermediate stages of this proposed process, all stages can be observed synchronically through an examination of the micro-variation across the language family. For example, Gikuyu uses the same element as an identificational copula and to convey predication focus, Haya uses the same element as an identificational copula and to mark progressive aspect, Kamba uses the same element to mark predicate focus and progressive aspect, whilst Gweno represents the most grammaticalized of the language types and uses the preverbal marker *ni* only to encode progressive (Philippson 2000).[18]

Comparative data from other East African Bantu languages which allow verb fronting is also taken as support for the use of variant word orders to encode pragmatic salience, with verb-auxiliary order and verb-fronting constructions seemingly in complementary distribution across the languages of the region.[19]

3.4 Support for the pathway of change: Meaning

3.4.1 The 'inversion contexts': support from information structure

Additional support for the pathway of change outlined in §3.2 comes from observations relating to word order variation in the languages from a synchronic perspective. While declarative main clauses in these six languages exhibit the verb-auxiliary order as outlined in this paper so far, this order is inverted (yielding auxiliary-verb order) in a number of syntactically conditioned contexts.

In Rangi, for example, this alternative order is found in future tense constructions that are associated with content interrogatives (68) sentential negation (69), clefts (70) and relative

[18] Thanks to a reviewer for alerting me to the case of Gweno.

[19] Such a proposal is further supported by the observation that there is dialectal variation in Simbiti with regards to the presence of the auxiliary-verb order (cf Section 2.8). Similarly, there are examples from Ngoreme which show that auxiliary-verb order is also possible (even outside of the expected syntactic contexts) with no apparent difference in interpretation (cf. example (36) above).

clauses (71). In contrast to declarative main clauses, post-verbal auxiliary placement is unacceptable in these contexts (72).

(68) *Ani á-rɪ wúl-a ma-papai a-ya?*
who SM1-AUX buy-FV 6-papaya DEM-6
'Who will buy these papayas?' (Rangi, Gibson 2012:114)

(69) *Nɪínɪ sí ndí-rɪ dóm-a na Kondoa tʊkʊ.*
1sg.PP NEG SM1sg-AUX go-FV CONN Kondoa NEG
'I will not go to Kondoa.' (Rangi, Gibson 2012:116)

(70) *Nɪ nɪínɪ ndí-rɪ kány-a ʊ-hʊ mʊ-ti.*
COP 1sg.PP SM1sg-AUX cut-FV DEM-3 3-tree
'It is me that will fell this tree.' (Rangi, Gibson 2012:227)

(71) *Mw-aarimʊ mw-eene á-rɪ loka a-boh-a*
1-teacher 1-REL SM1-AUX go-FV SM1-be.good-FV
'The teacher who will leave is good.' (Rangi, Gibson 2012:118)

(72) **Na nadi chw-a tú-rɪ vi-ryo?*
when harvest-FV SM1pl-AUX 8-millet
'When will we harvest millet?' (Rangi, Gibson 2012:115)

A similar situation is seen in Mbugwe, where the verb-auxiliary order is inverted in a number of contexts which closely resemble those identified in Rangi. These contexts include content questions (73), cleft constructions formed with *ne* (74) and relative clauses (75).[20]

(73) *nányu n-jé-mo-wéér-a á-ikal-ɛ neshopáki*
who SM1SG-FUT-1OM-tell-FV SM1-stay-SBJV national park
'Who will I tell to say in the national park?' (Mbugwe, Wilhelmsen 2018: 131)

(74) *Nɔ́ɔ́ng'ɔ́ ke-ká-vá ne n-síko ji-ɔ́ɔ́nsɛ vá-á-re j-a.*
so SM7-CONS-COP COP 10-day 10-all 2-PST-AUX.PROG1 come-FV
'And so it was that they were coming every day.'
Lit.: 'And so it was, it is every day they were coming.' (Mbugwe, Wilhelmsen 2018: 132)

(75) *Te-ré ke-ákora ke-ándá r-w-a ne va-nto.*
NEG-COP.LOC 7-food SM7-AUX.HAB1 eat-PASS-FV COP 2-person
'It was not food that was eaten by people.' (Mbugwe, Wilhelmsen 2018: 150)

[20] Wilhelmsen (2015) describes the constructions presented in (73)-(75) as simple constructions. As a result, what is considered in the current study to be the alternation between pre- and post-verbal auxiliary placement, is considered by Wilhelmsen (2015) to be an alternation between a simple and a periphrastic verb form. In the absence of independent evidence in support of analysing these elements as comprising a single 'word', the present account continues to consider such constructions as analogous to the compound auxiliary-based constructions seen throughout the Bantu family.

In Gusii, the past tense continuous form exhibits auxiliary-verb order in content questions (76), sentential negation (77) and relative clauses (78). This contrasts with the verb-auxiliary order found in declarative main clauses in the past tense continuous (cf. examples (22)-(27) above).

(76) *Ni-nki* *ki-aa-re* *ko-gwer-a* *e-nyoomba*?
COP-what SM7-CONT-AUX INF-fall-FV AUG-9.house
'What is falling on the house?' (Gusii, Evans Mecha p.c.)

(77) *Ti-o-ri* *ko-riing-or-a*
NEG-SM2sg-AUX 15-fold-RVRS-FV
'You are not unfolding.' (Gusii, Cammenga 2002: 386)

(78) *Maria* *n-eere* *o-re-enge* *ko-ges-a* *o-bo-ori*
Maria FOC-1.REL SM1-AUX-PAST INF-harvest-FV AUG-14-millet
'Maria is the one who is harvesting finger millet.' (Gusii, *author fieldnotes*)

On the basis of the current state of description the same alternation contexts also seem to apply for the other three languages – Ngoreme, Simbiti and Kuria. However, additional data would be required to determine whether the alternation contexts are restricted to exactly the same set of contexts. The contexts in which the auxiliary-verb order is found across these three languages can therefore be summarised as the following:

(79) The 'inversion' contexts
i) Content questions
ii) Negation
iii) Cleft constructions
iv) Relative clauses

All of the contexts in which the auxiliary-verb order is found are marked in terms of information structure. Wh-questions ask for new information, and wh-words are considered to be inherently focal. Accordingly, wh-questions and their answers are commonly used tests for focus (van der Wal 2016). Similarly, if affirmative constructions are considered to be unmarked then their negative counterparts can be considered to be marked in the sense that affirmative constructions can be considered more basic than negation constructions.[21] Cleft constructions and relative clauses are both used to indicate some backgrounded information against which the following proposition is assessed (Güldemann 2003). Clefted elements, for example, commonly convey the focus or topic of the clause (Hartmann & Veenstra 2011:20, Collins 1991) These functionally diverse construction types have also been noted to pattern together cross-linguistically (e.g. Schachter (1973), Drubig (2003), Takizala (1972)).

The presence of the verb-auxiliary/auxiliary-verb alternation, and specifically the contexts in which the alternation is found, provides additional evidence in support of the proposal that the verb-auxiliary order is related to information structure. The proposal developed here is that from a historical perspective, in the compound constructions associated with the standard auxiliary-verb word order, verb fronting was possible for pragmatic effects – to convey predication focus. However, in the environments outlined above, i.e. content questions, negation, cleft constructions, relative clauses, the information structure properties were

[21] The observation that the verb fronting construction described in varieties of Kikongo is also incompatible with negation is also relevant here (de Kind et al. 2015).

different from those in declarative main clauses. Since a focal reading was already conveyed on one or more of the given elements in these contexts, the constructions were not available for verb-fronting and so the grammaticalisation of the verb-auxiliary order did not extend to these constructions. This means that these so-called 'inversion' contexts actually represent the older, original word order that was historically found across all contexts.

This would also fit with the notion of reanalysis presented above and provide further support of the construction having its origin first in a verb fronting for predication focus purposes, and subsequently developing the association with progressive aspect. Such a directionality of change can be proposed on the basis that if the verb-auxiliary order was historically associated with predication focus, it would not have been available for already focal constructions such as content interrogatives or cleft constructions. However, the grammaticalization of information structure may mean that pragmatic relations lose their specificity (and iconicity) over time and that what were previous differences between pragmatic components become levelled out (Lehmann 2008).

This observation also fits with the broader notion of 'persistence' in the terms of Hopper (1991:28–30) whereby grammaticalizing markers tend to maintain certain features of their original sources. In the case of a variant word order used to convey predicate-centred focus, this has resulted in the verb fronting construction being incompatible with negative contexts, as is seen in all of the languages examined here, as well as with both backgrounding clauses (e.g. some clefts) and other focus types (e.g. interrogatives, relative clauses and some clefts).

One of the questions that remains at this stage is why the pragmatically-motivated word order variant would have developed into the standard (and indeed only) way of forming these constructions. Why did the verb-auxiliary order replace the auxiliary-verb order in certain contexts?. One possibility is that the construction filled a functional gap, perhaps if there was previously was no dedicated way of encoding progressivity, for example. Or perhaps there were other changes to the organisation of the tense-aspect system (or indeed the syntax as a whole) occurring at the same time which meant that this grammaticalization process became associated with progressive aspect and other strategies developed to encode focus. The presence of a relatively high number of second language speakers and high levels of multilingualism in the broader language ecology may have further served as a catalyst to the process of grammaticalisation (see §4).

3.4.2 A closer look at tense and aspect

The final observation in support of the proposal that the verb-auxiliary order has its origins in considerations of information structure comes from a closer examination of the tense-aspect distinctions with which this variant word order is associated.

In five of the six languages under examination in the current paper, the verb-auxiliary construction is found in the progressive, habitual and persistive aspects. The only language in which this generalisation does not hold is in Rangi where the verb-auxiliary order is associated with the future tense – a point that will be returned to in §3.5 below. A summary of the tense-aspect contexts in which the verb-auxiliary order is found in the six languages is shown in Table 2.

Table 2: Tense-aspect distribution of verb-auxiliary order in the languages

Language	**Tense-aspect distribution**

Rangi	Immediate future, general future
Mbugwe	Present imperfective, habitual, past progressive, future perfective
Gusii	Habitual, present continuous, past continuous, past habitual
Kuria	Present progressive
Simbiti	Present progressive, present habitual
Ngoreme	Present progressive, past progressive

Hyman and Watters (1984) and Güldemann (2003) examine evidence from a number of Bantu languages and propose that the progressive aspect is an inherently focal category in the language family. This claim is based on a number of observations. Firstly, in the progressive aspect, the ongoing nature of the event described by the verb constitutes the focus domain of the utterance (Güldemann 2003). Secondly, in many Bantu languages, the same morphological and syntactic devices are used for the expression of predicate focus and progressive meaning – either in the same language or from a cross-Bantu perspective (see §3.3.2). And finally, in more general terms, the progressive aspect views what the verb designates as an ongoing situation in which the referent of the verb is located. Verb focus constructions are therefore well suited to become grammaticalised into a progressive aspect (also noted by Lehmann (2008)).

The proposal made here is that the verb-auxiliary word order is an instantiation of the process of grammaticalisation identified by Güldemann (2003), which sees the development of forms which mark predicate focus into progressive interpretations. The specific tense-aspect combinations with which the verb-auxiliary is associated in these six languages further adds support to this proposal. In five of the six languages (all except Rangi) the verb-auxiliary order is associated with progressive aspect. In addition to progressive aspect, the order is also associated with habitual (in Gusii, Mbugwe and Simbiti), the imperfective (Mbugwe) and the future tense (Rangi and Mbugwe). In general terms, the progressive has been viewed as a specific meaning which often feeds into a chain of grammaticalization leading to more generalised meanings such as imperfective and present (Bybee et al. 1994 :127). As such, the progressive is considered to occur quite 'early' on in the process of grammaticalization, and may develop other subsequent meanings.

Cross-linguistic evidence shows constructions with progressive meanings commonly develop into presents or imperfects (e.g. in Turkic, Dravidian and Celtic languages Bybee et al. 1994:127). This means that the use of the verb-auxiliary order in the imperfective in Mbugwe for example fits within well-attested cross-linguistic patterns. Similarly, the progress can become a continuous by generalizing to apply to stative predicate as well as dynamic predicates (Bybee et al. 1994:139), meaning that the use of the verb-auxiliary order in Gusii can also be accounted for. Finally, the development of habituals from progressives – both of which occupy part of the imperfective spectrum – is also widely observed (Bybee et al. 1994:151).

All of the tense-aspect combinations with which the verb-auxiliary order is attested can therefore be attributed to widely observed cross-linguistic pathways of grammaticalization processes which have their origins, at varying degrees of remove, from the progressive. The

apparent exception to this generalisation is Rangi where the verb-auxiliary order is associated with the future tense. This is discussed below.

3.5 From predicate-centred focus to progressive to future tense

Rangi appears to stand out amongst the other languages and the account presented in §3.2 since the verb-auxiliary order is associated with the future tense rather than progressive aspect. However, based on cross-linguistically common pathways of grammaticalization, it is suggested here that this construction in Rangi is simply 'further grammaticalised' than in the other languages. This is based on the idea that the future tense interpretation is a natural extension of a previous system in which the verb-auxiliary order in Rangi was also associated with progressive aspect, as has been widely noted both in the Bantu languages and cross-linguistically.

In the Bantu languages, it has been noted that present and present progress forms often become near futures (Bastin 1989; Nurse 2008:118). In some varieties Kikongo, for example, the progressive in the verb fronting construction (discussed in §3.3.1) developed a tertiary future meaning out of its secondary progressive meaning. The suggestion is that this future marker has gained a greater independence from the verb fronting predicate-centred focus contexts than as a progressive marker. Although it has also been noted that a simple zero-tense-marked verb developed future time reference in the South Kikongo varieties Kisolongo, Kisikongo and Kizombo (Dom & Bostoen 2015), meaning that this future tense meaning may have arisen independently via analogy (de Kind et al. 2015:130).

From a cross-linguistic perspective, it is a well-documented that progressives have a tendency to further grammaticalize, often acquiring greater variability with regards to tense and eventually becoming imperfectives or general presents (see, for example, Bybee et al 1994:140–9). In such subsequent developments, progressives are likely to lose their pragmatic association with focus. Therefore Güldemann (2005:352) further describes progressives which are (still) associated with inherent focality as 'young' and having relatively recently grammaticalized, while those which are 'older' progressives tend to be no longer associated with focus, as is seen with the English progressive for example (Comrie 1976:33, Dahl 1985:90). Other cross-linguistic evidence in support of the development of the present progressive to the future tense can be seen in Romance languages such as Portuguese, Italian.

The proposal is that the immediate future construction in Rangi, which is formed using the auxiliary *-íise*, was the first verb-auxiliary construction to develop. The progressive aspect conveyed by this construction would have subsequently also been used to convey a near future reading – the reading with which it is most commonly associated in the present-day. The general future construction which employs the underspecified copula form *-rɪ* would then have developed by analogy with the immediate future form. This means that from a historical perspective, Rangi would also fit with the broader association of verb-auxiliary order with progressive aspects.

It is worth remembering however, that Gusii, Kuria, Ngoreme and Simbiti all also have verb-auxiliary constructions which use the simple copula form *-re*. This means that the possibility that the general future tense verb-auxiliary in Rangi developed independently or indeed, before the immediate future cannot be ruled out. However, the crucial characteristic of the Rangi general future tense construction is that there is no independent morphological marking of future tense in the general future tense which is constructed with *-rɪ*. This contrasts with the other auxiliary constructions which employ *-rɪ* but which also host specific tense-aspect information and the immediate future which alone employs the auxiliary *-íise*. It may well be the case therefore that the general future tense in Rangi developed to fill a

function gap since there was no dedicated future constructions, or indeed, that whatever the preceding future form was, had been lost.

4 Language contact and the development of verb-auxiliary order

4.1 Previous contact-based accounts of the verb-auxiliary order

Previous accounts of the verb-auxiliary order, at least in Rangi and Mbugwe, suggested that the construction might be the result of contact with non-Bantu languages (Mous 2000, Stegen 2002, Nurse 2003, Dunham 2005, Gibson 2013). The proposal for Rangi was based on the assumption that the verb-auxiliary order is not a Bantu inheritance, that it occurrs in highly restricted contexts and that Rangi has been in sustained contact with non-Bantu languages, specifically the Cushitic languages Burunge, Alagwa and Iraqw (additional details of the relevant linguistic ecology are provided below). The other four languages in which this order has been identified in the current paper are also spoken in areas of intense language contact. The question that arises is therefore whether the proposal that the verb-auxiliary order has its origins in processes of contact-induced change still hold now that a number of additional languages which exhibit this order have been identified?

4.2 A linguistic overview of the area

Rangi and Mbugwe are spoken in central Tanzania – an area known for its linguistic diversity with high degrees of multilingualism and language shift (Mous 2000, Kießling et al. 2008). Both Rangi and Mbugwe represent linguistic enclaves and are surrounded by speakers of non-Bantu languages. The nature of the language contact in the area is further characterised by the fact that, in addition to the Bantu languages, representatives of two other African language phyla are also found in the area. The Cushitic family is represented by the Iraqw, Gorwaa, Burunge and Alagwa, whilst the Nilotic languages found in the area are Datooga and Maasai. The non-Bantu language Sandawe is also present although its genetic affiliation remains the topic of debate. [22] The different language types represented by the linguistic families further impacts the nature of the language contact. In terms of basic word order for example, the Cushitic languages and Sandawe are SOV languages, the Bantu languages are typically SVO, whilst the Southern Nilotic language Datooga exhibits a predominantly VSO constituent order (Kießling et al. 2008). All linguistic observations of the Bantu and Cushitic languages in this area indicate that there has been significant interaction between the languages, and that they form a linguistic area or Sprachbund (Mous 2016).

Whilst the present-day speakers of Rangi and Mbugwe are separated by other linguistic groups, the two speech communities share a joint history and are thought to have moved into this region of Tanzania at the same time. The primary non-Bantu contact languages for present-day speakers of Rangi are the Cushitic languages Alagwa and Burunge, and to a lesser extent Gorwaa. There are a number of mixed Rangi/Alagwa and Rangi/Burunge villages throughout

[22] There is ongoing discussion regarding the genetic relation of the Tanzanian languages Sandawe and Hadza, and the Khoisan languages of Southern Africa – and indeed, whether the Khoisan languages of Southern Africa constitute a genetic grouping (Güldemann and Vossen 2000, Brenzinger 2013, Brenzinger 2014, Güldemann 2014). For support of the genetic connection between Sandawe and Hadza and the Khoisan family see Ehret (1986) and Elderkin (1986, 1989). For claims against the claim, the reader is referred to Westphal (1971:401) and Wright et al. (1995). For the purposes of the current study, the exact nature of the relationship is not central to the discussion. Rather, the case of Sandawe shows further linguistic diversity within the area.

the area (Kießling 2007:189) with high levels of bilingualism found particularly in Rangi by first language speakers of Alagwa and Burunge. It is worth noting however, that there is an asymmetry in this bilingualism, with a higher proportion of Alagwa- and Burunge-speakers speaking Rangi than the other way around.

The Rangi-speaking and Iraqw-speaking communities are no longer in direct contact, since Gorwaa is now spoken in the intervening area. However, on the basis of routes of migration and what is understood of the history of this region, Rangi and Iraqw are thought to have been in contact at some point in the past before Gorwaa split off and occupied the area in between these two languages (Kießling et al. 2008). The primary non-Bantu contact languages for Mbugwe are Gorwaa and Iraqw. Again, there is a sustained history of language contact between all three of these language groups (Mous 1993, 2000 Kießling et al. 2008).

There has also been contact between the Rangi/Mbugwe communities and speakers of Datooga (Nilotic), Sandawe (contested genetic affiliation) and Maasai. However, the nature of the interaction between the Rangi/Mbugwe communities and these other communities is less well documented and appears to differ from that found between the Rangi/Mbugwe and the Cushitic-speaking communities, perhaps due to the nature of their economies and lifestyles. The same level of bilingualism and interaction has not been noted amongst the Datooga-speaking communities, nor is there widespread the prevalence of 'mixed villages', at least in the case of the central Tanzania region. [23] Many more individuals who are of Alagwa or Burunge backgrounds are now part of the Rangi community, and raise their children speaking Rangi than those from other the communities (Oliver Stegen, p.c.).

The other four languages that exhibit the verb-auxiliary order – Gusii, Kuria, Ngoreme and Simbiti – are spoken in northern Tanzania and western Kenya in the area near Lake Victoria. These four languages are all in contact with non-Bantu contact languages to some extent, primarily the Nilotic languages Luo, Datooga and Maasai. Speakers of the Nilotic language Datooga have been present in the Mara region for over a thousand years and have lived alongside Bantu communities since then (Walker 2013:31). The Eastern Nilotic Maasai language is also present in the area and more recently, Western Nilotic Luo speakers have also moved to the region (Aunio 2015).

Only the languages Datooga and Maasai are common contact languages in both the central Tanzania region and the Lake region. As noted above, both of these languages play a more minor role as contact languages in the Central Tanzania region than for example, the Cushitic languages.

4.3 Challenges with a contact-based account of the verb-auxiliary order

Cross-linguistically, it has been noted that contact and borrowing may play a specific role in word order typology, with certain word orders only coming into existence as a result of contact. The word order in the Semitic language Amharic, for example, has been noted to have changed under influence from Cushitic (Leslau 1945, Harris & Campbell 1995: 124). Similar cases have been reported in Pipil, Xinca and Copainalá Zoque which are all described as having borrowed the VOS word order from neighbouring Mayan Languages (see Campbell 1987, Harris & Campbell 1995: 137).

In terms of a structural source of the marked verb-auxiliary word order, one possible source for its genesis in Rangi and/or Mbugwe could be found in Iraqw. Iraqw has a periphrastic future

[23] This differs from the contact between the Datooga-speaking communities and speakers of a number of Bantu language spoken in the Mara region of Tanzania which is the subject of a growing amount of research (see, e.g. Aunio 2015 for contact between Datooga and the Bantu language Ikoma also spoken in the Mara region and Roth 2014 for contact between Datooga and Ngoreme).

in which a verbal noun is followed by the auxiliary *aw* 'go', as can be seen in examples (80) and (81) below.

(80) *Makay i ma'á mahúngw ay-á.*
animals S.3 water:CON drinking:CON go:3-PL
'The animals will drink water.' (Iraw, Mous 1993:267)

(81) *Matlo atén a gadyée-r tleehhamá-r aw-aan-a-ká.*
tomorrow 1.PL S.1/2 work:CON-F doing:CON-F go-1.PL-INF-NEG
'Tomorrow we will not go to work.' (Iraqw, Mous 1993:267)

The proposal would be therefore that the verb-auxiliary order attested in the future tense in the Central Tanzania languages (i.e. Rangi and Mbugwe) is the result of structural transfer from Iraqw where an analogous construction involving a verbal noun and auxiliary (also with verb-auxiliary order) is used to encode the future tense. However, forwarding the periphrastic future constructions in Iraqw as the possible source for the borrowing of verb-auxiliary order into Rangi and Mbugwe is associated with a number of challenges.

The proposed analogous auxiliary *aw* 'go' in Iraqw could be seen as belonging to a much wider, cross-linguistically common pathway of change which sees the widespread use of verbs of motion, specifically those denoting concepts such as 'go', in the development of future tense auxiliaries (Heine 1993). Another challenge is that it has been noted that such constructions are marginal in Iraqw, and that a much more common strategy for the formation of auxiliary construction involves a verb expressing the meaning 'finish' which is used to convey the meaning 'to do something completely' (Maarten Mous, p.c. 2015).

There are also other aspects of the construction which do not support a contact-induced change analysis. The auxiliaries used in the formation of all of the compound constructions across the six languages appear to be of Bantu origin. As discussed in Section 2.9 the auxiliary *-rɪ* – and its variants – can be assumed to be of Bantu inheritance (most likely related to the reconstructed Proto-Bantu copula *-*dì* (Guthrie 1967-71 Vol. III: 150, C.S. 547; BLR3 940) or *-*di* (Meeussen 1967: 86)). Similarly, in Mbugwe, auxiliary forms in many instances appear to be transparently linked to main verbs (a common source of auxiliaries across Bantu (Botne 1989: 169)). For example, the auxiliary *-je* appears to be derived from the verb *-ja* 'to come' and the auxiliary *kééndé/kééndé* from the Proto-Bantu form **gend* 'walk, travel, go, go away' (Wilhelmsen 2018). Similarly, in Rangi, the auxiliary *-íise* may have its origins in a now obsolete, verb form based on PB *-*yìj* 'come' (cf. Guthrie 1967-71 Vol. IV: 176, C.S. 2045; BLR3 3425).

This would mean that a contact-induced account of the verb-auxiliary order in these languages would represent an instance of borrowing of structure but not of form. Thus, the auxiliaries would be a Bantu inheritance, but have been imposed onto a non-Bantu structure. This may have meant that second-languages speakers of Rangi, for example, with a Cushitic first language would have used Rangi lexical features with the relevant structural features from their first language (cf. Thomason and Kaufman 1988, Van Coetsem 1988, Guy 1990).

If the verb-auxiliary order were indeed an instance of structural borrowing, it might also be expected that there would be other examples of borrowing and interference identifiable in the languages. For example, it would be expected for there to be a greater degree of evidence of lexical borrowing. This would be in line with the approach forwarded by Thomason & Kaufmann (1988) who propose a 'borrowing scale' which articulates types of contact-induced change and the levels of contact with which they are commonly associated. For structure to be borrowed and to result in changes in word order, for example, is considered to occur only in

cases of intense contact. While there is evidence of such contact, there is only limited evidence of transfer and interference Roth (2014, 2018) and Aunio (2015), Gibson & Marten 2019 at the more basic levels of lexicon and function words, which is assumed to come before the transfer of structural input. As such, the extent of this interference does not conclusively support a contact-based account of the rise of the verb-auxiliary order

5 Summary and concluding remarks

This paper has presented a discussion of the verb-auxiliary order found in a subset of East African Bantu languages with a focus on the processes of change which are proposed to have been responsible for its genesis. The verb-auxiliary order in these languages is unusual in the context of East African Bantu, as well as in the wider context of SVO languages. While the presence of verb-auxiliary order has previously been noted in Rangi (Dunham 2005, Stegen 2002, Gibson 2012), Mbugwe (Mous 2000, 2004, Wilhelmsen 2014), Gusii (Kerr 2017), Simbiti (Walker 2013) this paper constitutes the first systematic study of this word order in all six languages and from a broader comparative perspective.

All six languages are spoken in areas of high linguistic diversity and have been in sustained contact with non-Bantu languages – primarily Cushitic languages in the case of Rangi and Mbugwe, and Nilotic languages in the case of Gusii, Kuria, Simbiti and Ngoreme. Despite the presence of the verb-auxiliary order, these languages exhibit otherwise Bantu-typical head-initial syntax and the morphosyntax commonly associated with East African Bantu languages. Similarly, there appears to be nothing in particular which sets these languages apart from other Bantu languages, many of which are also in contact with non-Bantu languages but do not exhibit verb-auxiliary order.

Previous studies have proposed a contact-induced account for the origins of the verb-auxiliary order, at least for Rangi and Mbugwe (Mous 2000, Stegen 2002, Dunham 2005, Gibson 2013), albeit without providing details of the structural source of this process of transfer or borrowing. However, here it is claimed that the presence of high levels of bi- and multi-lingualism in Bantu and non-Bantu languages is not enough to conclude that the verb-auxiliary order is purely a contact feature. Rather, the present study proposes a grammaticalisation account of the origins of the marked word order on the basis of language-internal and comparative evidence. In all six languages, the constructions that exhibit verb-auxiliary order encode progressive, habitual or persistive aspects (or future tense in the case of Rangi). Similarly, whilst declarative main clauses exhibit verb-auxiliary order, the auxiliary-verb order is found in syntactically-conditioned contexts all of which are marked in information structural terms, namely wh-questions, sentential negation, relative clauses and cleft constructions. These observations are taken together as evidence that the verb-auxiliary order found in these languages has its historical origins in pragmatic considerations under which the verb could appear before the auxiliary in order to encode predication focus. A subsequent stage of grammaticalization saw the development of this marker of predication focus into progressive aspect. In some of the languages, this was followed by subsequent stages of grammaticalization into other aspectual or temporal domains. The pathway of change therefore further corroborates the broader claim that the present progressive is an inherently focal category in Bantu (Hyman & Watters 1984, and Güldemann 2003) and represents another example of the polysemy between markers of focus and progressive aspect.

However, it may well be the case that the high levels of language contact and bilingualism found in the communities that speak these six languages served as a catalyst for the process of language change documented by the rise of the verb-auxiliary order. In this sense, the account developed here fits with the proposal that even processes of contact-induced language change also involve stages of grammaticalization (Heine & Kuteva 2005). The account developed here

also fits with the proposal that information structure is part of grammatical structure and that, as a result, may come about through grammaticalization and may dissolve by grammaticalization (Lehmann 2008: 207).

References

Abels, Klaus and Peter Muriungi. 2008. The focus marker in Kîîtharaka: Syntax and semantics. *Lingua*. 687–731.

Anonymous. n.d. 'Suba-Simbiti survey: Additional linguistic data comparison of two Simbiti varieties'. SIL. Unpublished manuscript.

Ashton, Ethel O. 1947. *Swahili grammar, including intonation*. Longman.

Aunio, Lotta. 2015. 'Contact leading to a more complex tonal system: The case of Ikoma and Datooga in north-west Tanzania'. Unpublished ms. University of Helsinki.

Aunio, Lotta. 2015. A typological perspective on Bantu nominal tone: the case of Ikoma-Nata-Isenye in western Tanzania'. *Southern African Linguistics and Applied Language Studies* 2015, 33(3), 359–371.

Aunio, Lotta, Holly Robinson, Tim Roth, Oliver Stegen and John B. Walker. 2019. 'The Mara languages (JE40)'. In: *The Bantu Languages*. Koen Bostoen, Derek Nurse, Gérard Philippson and Mark van de Velde, (eds). 2nd edition. Routledge.

Bastin, Yvonne. 1989. E'l prefijo locativo de la clase 18 y la expresión del progresivo presente en bantú (I)'. Estudios Africanos. *Revista de la Asociación Española de Africanistas*. 4,7:61–86

Bennett, Patrick R. 1985. 'Gikuyu ni Kioigire: essays, texts and glossaries'. Unpublished ms. The African Studies Program, University of Wisconsin-Madison.

Bostoen, Koen & Gilles-Maurice de Schryver. 2015. Linguistic innovation, political centralization and economic integration in the Kongo kingdom: Reconstructing the spread of prefix reduction. *Diachronica* 32. 139–185.

Bostoen, Koen and Sebastian Dom. 2015. 'Examining variation in the expression of tense/aspect to classify the Kikongo Language Cluster' *Africana Linguistica*. 21, 163–211.

Bouka, L.Y. 1989. 'Elements de description du kaamba. Parler bantou de la Republique Populaire du Congo (Groupe Koongo, HI7b)'. Unpublished ms. Universite Libre de Bruxelles.

Botne, Robert. 'Specificity in Lusaamia infinitives.' *Studies in Language.* 28.1 (2004): 137–164.

Brenzinger, Matthias. 2013. 'The twelve modern Khoisan languages'. Paper presented at Khoisan Languages and Linguistics, Proceedings of the 3rd International Symposium (Research in Khoisan Studies 29), Cologne: Rüdiger Köppe.

Brenzinger, Matthias. 2014. 'Classifying the non-Bantu click languages'. In: *Papers from the Pre-Colonial Catalytic Project (Volume 1)*. Lungisile Ntsebeza and Chris Sanders, (eds). Cape Town: University of Cape Town. 80–102.

Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The evolution of grammar*. Chicago & London: The University of Chicago Press.

Cammenga, Jelle. 2002. *Phonology and Morphology of Ekegusii: A Bantu language of Kenya*. Köln, Rüdiger Köppe Verlag.

Cammenga, Jelle. 2004. *Igikuria Phonology and Morphology: A Bantu language of South-West Kenya and North-West Tanzania*. East-African languages and dialects. Köln, Rüdiger Köppe Verlag.

Campbell, Lyle. 1987. Syntactic change in Pipil. *International journal of American Linguistics*. 53. 253–80.

Collins, Peter. 1991. *Cleft and Pseudo-Cleft Constructions in English*. London. Routledge.

Dahl, Östen. 1985. Tense and Aspect Systems. New York. Basil Blackwell.

de Kind, Jasper, Sebastian Dom, Gilles-Maurice de Schryver and Koen Bostoen. 2015. 'Event-centrality and the pragmatics-semantics interface in Kikongo: From

predication focus to progressive aspect and vice versa.' *Folia Linguistics Historica*(36): 113–163.
Dik, Simon. C. 1997. *The Theory of Functional Grammar, Part 1: The structure of the* clause. Berlin: Mouton de Gruyter.
Dom, Sebastian & Koen Bostoen. 2015. Examining variation in the expression of tense/ aspect to classify the Kikongo language cluster. *Africana Linguistica* 21. 163–211.
Drubig, Hans Bernhard. 2003. 'Toward a typology of focus and focus constructions.' *Linguistics* 41(1): 1–50.
Dryer, Matthew S. 1992. 'The Greenbergian word order correlations'. *Language*. 81–138.
Dunham, Margaret. 2005. *Eléments de description du langi, langue Bantu F.33 de Tanzanie*. Collection Langues et littératures de l'Afrique noire SELAF 413. Louvain-Paris-Dudley MA, Peeters.
Ehret, Christopher. 1986. 'Proposals on Khoisan reconstruction.' *Sprache und Geschichte in Afrika* 7(2): 105–130.
Elderkin, Edward D. 1986. 'Diachronic Inferences from basic sentence and noun Structure in Central Khoisan and Sandawe' *Sprache und Geschichte in Afrika* 7: 131–156.
Elderkin, Edward D. 1989. *The significance and origin of the use of pitch in Sandawe*. PhD dissertation. Department of Language and Linguistic Science, University of York. .
Gibson, Hannah. 2012. *Auxiliary placement in Rangi: A Dynamic Syntax perspective*. PhD dissertation. SOAS University of London.
Gibson, Hannah. 2013. 'Auxiliary placement in Rangi: A case of contact-induced change?' *SOAS Working Papers in Linguistics* 16: 153–166.
Gibson, Hannah. 2016. 'A unified dynamic account of auxiliary placement in Rangi'. *Lingua*.184. 79–103.
Gibson, Hannah. 2019. 'Auxiliary constructions in Bantu' Unpublished manuscript, University of Essex.
Gibson, Hannah and Lutz Marten. 2019. 'Probing the interaction of language contact and internal innovation: four case studies of morphosyntactic change in Rangi'. *Studies in African Linguistics*, 49.
Greenberg, Joseph H. 1966. 'Some universals of grammar with particular reference to the order of meaningful elements'. In: *Universals of Language*. Joseph H. Greenberg, (ed). Cambridge, Massachusets and London, England, MIT Press**:** 73–113.
Güldemann, Tom. 2003. 'Present progressive vis-à-vis predication focus in Bantu: A verbal category between semantics and pragmatics.' *Studies in Language* 27: 323–360.
Güldemann, Tom. 2014. 'Khoisan' language classification today'. In: *Beyond 'Khoisan': Historical Relations in the Kalahari Basin*. Tom Güldemann and A.-M. Feh, (eds). Amsterdam/Philadelphia, John Benjamins: 1–41.
Güldemann, Tom, Ines Fiedler and Yukiko Morimoto. 2014. 'The verb in the preverbal domain across Bantu: infinitive "fronting" and predicate-centered focus'. Unpublished ms. Unpublished ms. Humboldt University Berlin.
Gunnink, Hilde. 2016. 'The fronted-infinitive in Fwe' Paper presented at 6th International Bantu conference, Helsinki University. 20-23rd June 2016.
Guthrie, Malcolm. 1967–71. *Comparative Bantu*. Farnborough, Gregg.
Guy, Gregory. 1990. 'The sociolinguistic types of language change.' Diachronica 7(1): 38-47.
Hadermann, Pascale. 1996. 'Grammaticalisation de la structure infinitive+verbé conjugue dans quelques langues bantoues.' *Studies in African Linguistics* 25(2): 155–169.
Harris, Alice & Lyle Campbell. 1995. Historical syntax in cross-linguistic perspective. Cambridge: Cambridge University Press.

Hartmann, Katharina and Tonjes Veenstra. 2013 'Introduction to Cleft structures' In: Hartmann & Veenstra (eds). *Cleft Structures* Veenstra. 1–32.

Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization* Oxford, Oxford University Press.

Heine, Bernd, and Mechtild Reh. 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Helmut Buske.

Heine, Bernd, and Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27 (3), 529 –572.

Heine, Bernd, and Tania Kuteva. 2008. 'Constraints on contact-induced linguistic change'. *Journal of Language Contact.* 2(1), 57–90.

Hopper, Paul J. 1991. On some principles of grammaticalization. IN Elizabeth Closs Traugott & Bernd Heine (Eds.) *Approaches to* Grammaticalization. 17 –35. Amsterdam: Benjamins.

Hyman, Larry and John Watters. 1984. 'Auxiliary focus.' *Studies in African Linguistics* 15: 233–273.

Kerr, Elisabeth J. 2017. *Focus marking in Ekegusii*. BA dissertation. Department of Theoretical and Applied Linguistics, University of Cambridge.

Kießling, Roland. 2007. 'The "marked nominative" in Datooga.' *Journal of African Languages and Linguistics* 28 (2): 149–191.

Kießling, Roland, Maarten Mous and Derek Nurse. 2008. 'The Tanzanian Rift Valley area. A Linguistic Geography of Africa'. In: *A linguistic geography of Africa*. Bernd Heine and Derek Nurse, (eds). Cambridge, Cambridge University Press**:** 186–227.

Kuteva, Tania. 2008. 'On the frills of grammaticalization'. In: María José López-Couso, Elena Seoane & Teresa Fanego (eds.), *Rethinking grammaticalization: New perspectives*. Amsterdam a Philadelphia: John Benjamins. 189–217.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. *Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lehmann, Christian. 2008. 'Information structure and grammaticalization'. In Elena Seoane and María José López-Couso (eds.) *Theoretical and Empirical Issues in Grammaticalization*. Typological Studies in Language 77. 207–229.

Leslau, Wolf. 1945. 'The influence of Cushitic on the Semitic languages of Ethiopia: a problem of substratum'. *Word*.1. 59-82.

Marten, Lutz. *To appear*. Bantu and Bantoid. In Rainer Vossen (ed.), Handbook of African Linguistics. Oxford: Oxford University Press.

Marten, Lutz, Peter Edelesten, Rozenn Guérois and Hannah Gibson. 2018. Database of morphosyntactic variation in Bantu. Leverhulme Trust funded project 'Morphosyntactic variation in Bantu: typology, contact and change'.

Maho, Jouni Filip.1999. *A comparative study of Bantu noun classes*. Acta Universitatis Gothoburgensis. Gothenburg, Sweden.

Meeussen, Achille Emille. 1967. *Bantu Grammatical Reconstructions*. Africana Linguistica. Tervuren.

Morimoto, Yukiko. 2013. 'Gĩkũyũ'. Unpublished ms. Paper presented to the Predicate-centered focus types: B7 project meeting, Humboldt-Universität zu Berlin, Nov. 19, 2013.

Morimoto, Yukiko. 2016. 'Verb doubling vs. the conjoint/disjoint alternation' Paper presented at 6th International Bantu conference, Helsinki University 20-23rd June 2016.

Morimoto, Yukiko. 2017. 'The Kikuyu focus marker nĩ: formal and functional similarities to

the conjoint/disjoint alternation'. In van der Wal, Jenneke and Larry M. Hyman (eds.), *The conjoint/disjoint alternation in Bantu*, 147–174. Berlin: De Gruyter.

Moshi, Lioba. 1998. 'Word order in multiple object constructions in KiVunjo-Chaga.' *Journal of African Languages and Linguistics* 19: 137–152.

Mous, Maarten. 1993. *A grammar of Iraqw*. Cushitic Language Studies, 9. Hamburg, Helmut Buske.

Mous, Maarten. 2000. 'Counter-universal rise of infinitive-auxiliary order in Mbugwe (Tanzania, Bantu, F34.)'. In: *Mehr als nur Worte... : afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland* Rainer Vossen, Angelika Mietzner, Antje Meissner, (ed). Köln, Köppe**:** 469–481.

Mous, Maarten. 2004. *A grammatical sketch of Mbugwe: Bantu F34, Tanzania*. Grammatische Analysen afrikanischer Sprachen. Köln, Rüdiger Köppe Verlag.

Mous, Maarten. 2016. *Alagwa Grammar, Texts and Lexicon*. Cushitic Language Series. Köln, Rüdiger Köppe Verlag.

Ndumbu, J. M. G. and W. H. Whiteley. 1962. 'Some problems of stability and emphasis in Kamba one-word tenses.' *Journal of African Languages* 1(2): 167–180.

Nurse, Derek. 2000. 'Diachronic morphosyntactic change in Western Tanzania'. In: *Mehr als nur Worte... : afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland*. Rainer Vossen, Angelika Mietzner and Antje Meissner, (eds). Köln, Köppe**:** 517–534.

Nurse, Derek. 2003. 'Aspect and tense in Bantu languages'. In: *The Bantu Languages*. Derek and Gérard Philippson (eds.) London, Routledge**:** 90–102.

Nurse, Derek. 2008. *Tense and Aspect in Bantu*. Oxford, Oxford University Press.

Nurse, Derek & Gérard Philippson. 2006. 'Common TA markers in Bantu' *Journal of African Languages and Linguistics*. 27. 155–196.

Purvis, John Bremner. 1907. *A manual of Lumasaba Grammar*. Society for Promoting Christian Knowledge.

Riedel, Kristina. 2009. *The syntax of object marking in Sambaa*: *A comparative Bantu perspective*. Netherlands Graduate School of Linguistics.

Roth, Tim. 2014. 'The Case of Ngoreme: Distinguishing contact from inheritance in Mara Bantu'. Paper presented at Language Contact: State of the Art, Helsinki, Finland.

Roth, Tim. 2018. *Aspect in Ikoma and Ngroreme: a comparison and analysis of two Western Serengeti Bantu languages*. Phd dissertation. University of Helsinki.

Schachter, Paul. 1973. 'Focus and relativization' *Language*: 19–46.

Schwarz, Florian. 2003. *Focus marking in Kikuyu*. MA dissertation. Humboldt-Universität zu Berlin.

Schwarz, Florian. 2003. 'Focus marking in Kikuyu.' *ZAS Papers in Linguistics* 30: 41–118.

Schwarz, Florian. 2007. 'Ex-situ focus in Kikuyu'. In: *Focus strategies in African languages: the interaction of focus and grammar in Niger-Congo and Afro-Asiatic*. Enoch O. Aboh, Katharina Hartmann and Malte Zimmermann, (eds). Berlin, Mouton de Gruyter**:** 139–160.

Stegen, Oliver. 2002. 'Derivational processes in Rangi.' *Studies in African Linguistics*. 31(1/2): 129–153.

Stegen, Oliver. 2006. Participatory Research, TAM Functions, and Dialect Survey Among the Rangi. SIL, unpublished manuscript.

Takizala, Alexis. 1972. 'Focus and relativisation: The case of Kihung'an' *Studies in African Linguistics*. 3(2). 259–286.

Thomason, Sarah G. 2001. Contact-Induced Language Change and Pidgin/Creole Genesis. In: Norval Smith & Tonjes Veenstra (eds.), *Creolization and contact* Amsterdam & Philadelphia: John Benjamins. 249–262.

Van Coetsem, Frans. 1988. Loan phonology and the two transfer types in language contact. Dordrecht: Foris.

van de Velde, Mark & Koen Bostoen. 2019. *The Bantu languages*. 2nd edition. London: Routledge.

van der Wal, Jenneke. 2016. 'Diagnosing focus.' *Studies in Language* 40(2): 259–301.

Walker, John B. 2013. *Comparative tense and aspect in the Mara Bantu languages: Towards a linguistic history*. MA dissertation. Trinity Western University.

Westphal, Ernst O. J. 1971. 'The click languages of southern and eastern Africa'. In: *Current trends in linguistics 7: Linguistics in Sub-Saharan Africa*. Thomas A. Sebeok, (ed). The Hague, Mouton**:** 367–420.

Whiteley, Wilfred. H. 1959. *The tense system of Gusii*. Kampala, East African Literature Bureau.

Wilhelmsen, Vera. 2014. 'Periphrastic verbs in Mbugwe (F34)'. Unpublished ms. Uppsala University.

Wilhelmsen, Vera. 2018. *A description of the Mbugwe lanaguage – with a focus on tone and TAM*. Phd dissertation. Uppsala University.

Wright, Richard, Ian Maddieson, Peter Ladefoged and Bonny Sands. 1995. 'A phonetic study of Sandawe clicks.' *UCLA Working Papers in Phonetics* 91: 1–24