Andrew Harvey*, Hannah Gibson and Richard Griscom

Preverbal clitic clusters in the Tanzanian Rift Valley revisited

Abstract: This paper examines preverbal clitic clusters in the Tanzanian Rift Valley, an area of high linguistic diversity with representatives of the Bantu, Cushitic, and Nilotic families, as well as Sandawe (possibly a distant member of the Khoi-Kwadi family), and the language isolate Hadza. An earlier work (Kießling, Roland, Maarten Mous & Derek Nurse. 2008. The Tanzanian Rift Valley area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 186–227. Cambridge: Cambridge University Press) identified preverbal clitic clusters as a widespread feature across many languages of the Rift Valley, and posited the preverbal clitic cluster as a feature characteristic of a 'Tanzanian Rift Valley Area'. The current paper provides further detail on preverbal clitic clusters across the languages of the region and examines possible routes of development for these structures. From this analysis, the picture that emerges is complex: contact scenarios cannot be restricted to ones in which West Rift Cushitic or its predecessor languages are the only models for the development of a preverbal clitic cluster and, in the case of Sandawe (and perhaps the Datooga varieties), it appears as if the development of a preverbal clitic cluster cannot be linked to contact at all. In terms of what this means for the 'areality' of the Tanzanian Rift Valley, this paper forgoes discussions about geographical delineation or arguments for or against a 'Tanzanian Rift Valley Area' in favour of highlighting the individual historical events (c.f. Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press) that may have given rise to preverbal clitic clusters in the languages of our sample, as well as encouraging continued investigation into the nature of these histories, both from a linguistic and interdisciplinary perspective.

*Corresponding author: Andrew Harvey, University of Bayreuth, Bayreuth, Germany,
E-mail: andrewdtharvey@gmail.com. https://orcid.org/0000-0003-3659-4704
Hannah Gibson, University of Essex, Colchester, UK, E-mail: h.gibson@essex.ac.uk. https://orcid.org/0000-0003-2324-3147
Richard Griscom, Independent Researcher, Oregon, USA, E-mail: rgriscom@gmail.com. https://orcid.org/0000-0003-2150-7065

Mukhtasari wa Kiswahili [Swahili Abstract]: Chapisho hiki huangalia vikundi vya viangami vinavyokaa kabla ya kitenzi kwenye Bonde la Ufa la Tanzania, eneo lenye anuwai ya lugha kubwa sana, na lugha kutoka familia za Kibantu, Kikushi, Kiniloti, pamoja na Kisandawe (kinachoweza kuwa na mnasaba na lugha za familia ya Khoi-Kwadi), na lugha kisiwa ya Kihadzabe. Kazi mmoja ya mapema zaidi (Kießling, Roland, Maarten Mous & Derek Nurse. 2007. The Tanzanian Rift Valley area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 186–227. Cambridge: Cambridge University Press) ilivitambua vikundi vya viangami vinavyokaa kabla ya kitenzi kama sifa bainifu kwenye lugha nyingi za "Eneo la Bonde la Ufa la Tanzania". Chapisho hiki hutoa taarifa zaidi kuhusu vikundi vya viangami vinavyokaa kabla ya kitenzi kwenye lugha za eneo hili, pamoja na kuchunguza njia za ukuzaji wa viunzi hivi. Kutokana na uchambuzi huu, picha inayopatikana ni yenye changamani: mipangilio ya matukio hayawezi kuweka Kikushi cha Ufa Magharibi (yaani: West Rift Cushitic) au lugha zake tangulizi kama mtindo kwa ukuzaji pekee wa vikundi vya viangami vinavyokaa kabla ya kitenzi, na, kwa Kisandawe (na labda lugha za Kidatooga), inaonekana kwamba ukuzaji wa vikundi vya viangami vinavyokaa kabla ya kitenzi hauwezi hata kuelezwa kama tokeo ya mgusano wa lugha. Kuhusu uwezo wa Bonde la Ufa la Tanzania kuwa "kundi eneo", chapisho hiki huacha mazungumzo wa jiografia na hoja kwa au dhidi ya "Kundi Eneo la Bonde la Ufa la Tanzania" na huongea zaidi kuhusu matukio moja moja ya kihistoria (linganisha Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press) yaliyoweza kusaidia ukuzaji wa vikundi vya viangami vinavyokaa kabla ya kitenzi kwenye lugha sampuli yetu, na vile vile hutumainisha uchunguzi zaidi kwenye historia hizi kwa vifaa vya isimu lakini vile vile za masomo mbalimbali.

1 Introduction

This paper examines clitic clusters in the Tanzanian Rift Valley (TRV). The Tanzanian Rift Valley is a region of high linguistic diversity with representatives of the Bantu, Cushitic and Nilotic language families, as well as Sandawe – possibly a distant member of the Khoe-Kwadi family – and the language isolate Hadza. A seminal chapter by Kießling, Mous, and Nurse (hereafter: (KMN 2008)) identified a concentration of preverbal clitic clusters in the Tanzanian Rift Valley Area. Crucially, they noted that preverbal clitic placement is neither a prominent feature of Nilotic nor the Bantu languages of the area and as such, they considered preverbal clitic clusters to represent an areal feature reflecting the sustained history of language contact in the

area. The prevalence of preverbal clitic clusters was only one of a number of features identified by KMN (2008) as characteristic of the linguistic area. The present paper builds on the foundation laid out by KMN (2008) but expands on this to include a broader range of languages and allows for more detailed discussion and the inclusion of additional data through the development of a paper dedicated to clitic clusters in the TRV.

The paper examines the presence of these clitic clusters in the region and seeks to interrogate both the category and the distribution of the forms. For the notion of 'clitic cluster' we draw particularly from KMN (2008). This draws on the descriptive work on the topic, and reflects structures exhibited by a range of languages which share a number of core features. Mous (2001: 125) describes these as "verbal functions [which are] divided over the verb and an obligatory sentence building word." This has been noted to be a widespread feature of Proto-West Rift, as well as a number of its present-day successor languages, many of which have a "distributed predicative syntax" (KMN 2008: 197).

The present paper takes the languages which were included in the KMN (2008) study as a starting point and seeks to explore this particular feature, captured therein under the heading 'preverbal clitic cluster' in more detail. The present study draws on data from 12 languages, including Cushitic, Nilotic and Bantu languages, as well as Sandawe and Hadza. These languages reflect distinct language families (and relatedly, a diversity of structures) represented by the languages of the Rift Valley Area. In terms of structures, we have taken the approach adopted by KMN (2008) as a starting point and include both auxiliary-type constructions (as found in the Bantu languages Nyaturu, Rangi and Mbugwe), as well as those of the South Cushitic type of preverbal clitic clusters. With an improved descriptive status for the languages under examination, and the inclusion of additional languages, the present study contributes to our knowledge of the languages in the Area, the distribution of the pre-verbal clitic category, as well as furthering our understanding of the prevalence of clitic clusters as an areal feature in the Rift Valley.

As will be seen over the course of the paper, there are some languages for which there is clearer evidence for this preverbal clitic cluster, or for which the existence of such a structure is more readily available. However, we argue here that the concentration of features encoding concepts related to tense-aspect-mood (as well as, where relevant, clause type etc.) in a particular location within the clause motivates and justifies a united approach across these diverse languages. The paper seeks to provide not only an update on the study conducted by KMN (2008) in that the present paper incorporates further data which were not available at the time of publication – but aims also to expand both the empirical coverage and the associated discussion through the development of a paper where clitic clusters constitute our sole focus.

The paper is structured as follows: Section 2 provides an overview of the languages of the study, detailing relevant socio-historical background, contact languages and typological and descriptive background on the languages under examination. Section 3 provides an in-depth exploration of the pre-verbal clitic clusters (PCC) in the languages of the study. Section 4 explores the possible origins of these clitic clusters, in light of the prevalence of this construction type in the Tanzanian Rift Valley and revisits the analysis presented by KMN (2008) in terms of the possible pathways of development that may have given rise to the structures. Section 5 offers some concise conclusions and highlights directions for future work.

Finally, a note on the presentation of data is in order. Data presented here represent a combination of our own original field-based data collection, data generously shared with us from a range of different sources, and published sources. Data are presented in a four-line format, where the first line (italicised) represents the utterance as-encountered in its original source. The second line presents a parsed version of the utterance (in bold). The third line provides the glosses of the corresponding morphs in the parsed line directly above. The fourth line is a free translation of the utterance. For the most part, all data presented here are shown as in the original source and where relevant this is supplemented with discussion with the authors and/or with native speakers. One major change is that, where an original source does not identify a morpheme as a clitic which we do identify as a clitic, we have modified the original to indicate this (i.e. by adding the = sign).

2 Clitics in the Rift Valley area

2.1 Context of the study

KMN (2008) "The Tanzanian Rift Valley Area", a chapter in the larger title *A Linguistic Geography of Africa* (Heine and Nurse eds.) is the first work to examine preverbal clitic complexes of the TRV together. In broad terms, the chapter section (pages 197–206) observes that, although preverbal clitic complexes of this type [i.e. those which have extended the complex through fusion with conjunctions and adpositions etc.] are characteristic of Southern Cushitic, they are not typical in either Bantu or Nilotic. However, within the TRV, both Bantu (esp. Nyaturu) and Nilotic (esp. Gisamjanga Datooga) languages "display incipient preverbal clitic clustering" (page 198). From a historical viewpoint (page 199), preverbal clitic complexes in both Nyaturu and Gisamjanga Datooga are said to have arisen as a result of long-term contact between speakers of pre-Nyaturu and speakers of pre-Gisamjanga with speakers of Proto West Rift (the hypothetical ancestor of the current South Cushitic languages) (page 199). These are observations from KMN's (2008) paper to which we will return over the

course of the paper as we explore these constructions in more detail. The Bantu languages Nyilamba, as well as Kimbu (though the data for this latter language is sparse) do have preverbal clitic clusters, but not to the same degree as Nyaturu (page 201). The same historical dynamic as proposed for Nyaturu is also proposed for these two languages (page 199). Sandawe (possibly either a language isolate or a distant member of Khoe-Kwadi) has a series of person-gender-number (PGN) markers, which are taken as a close analogue of preverbal clitic complexes (page 205). Historically (and because of similarities with forms in Khoe-Kwadi), it is judged that the Sandawe and South Cushitic forms are due to chance rather than contact, though the two language groups being in contact may have reinforced the structures present in each (page 205). The Hadza language (isolate) is seen as having similar structures (page 203), but it was decided that the available data were not sufficient for further analysis (page 206). The other languages in the sample, Rangi and Mbugwe, are judged to not have preverbal clitic complexes, and are not discussed in the section.

KMN (2008) is the model on which we have based this paper, as such, we will be considering all the languages included in their chapter. We also focus on the structures identified therein as PCCs. Following this brief summary, a few statements can now be made about the limits of KMN (2008), and how this paper will respond to them.

First, and as a result of the fact that their work could only dedicate one sub-section of one book chapter to the topic, examples of the preverbal clitic clusters of each language of the area were not provided. In the present paper, we have space to provide examples and analysis of examples for every language for which we have data. Second, KMN (2008) takes the South Cushitic PCC to be the prototype for ex-aminations of the structures under assessment. Under this assumption, South Cushitic is taken both to be the 'source' of the prevalence and concentration of this structure in the region, and to constitute the structure on which the criteria for what constitutes PCC is based. The present paper critically examines this position, giving a finer-grained examination of the preverbal clitic clusters mentioned, both in terms of their similarities and differences to the South Cushitic prototype, as well as their proposed historical origins. The picture that emerges shows a contact situation which is rather more complex. Third, several languages are considered homoge-neous in terms of how their PCCs behave.[1] The South Cushitic languages (Alagwa, Burunge, Gorwaa, and Iraqw) are all considered as a single group. Nyilamba and Kimbu are considered another group. Rangi and Mbugwe are considered another,

---

1 In reality, this paper will only go part of the way in demonstrating the full linguistic diversity of the region, as most of the languages listed here are composed of more or less distinct varieties, many of which remain undescribed and otherwise unrepresented in the literature. Where distinct varieties of a language exist, they will be mentioned in Section 2.3, with a note given on the variety used in this paper.

and the Datooga varieties yet another. The present paper examines these assumed groupings, and indicates where forms differ.

Since the publication of KMN (2008), Anderson (2011) has also commented on the preverbal clitic complexes of the TRV, this time from the point of view of auxiliary verb constructions (AVCs). Anderson does not aim to demonstrate that the constructions under discussion constitute AVCs, but rather operates under the assumption that they do, and uses a typology of AVCs to compare the constructions based on their morphosyntactic and semantic features. Anderson finds no areal AVC profile for the TRV and reports that many of the languages reflect the typological profiles of their genetic groupings. Anderson proposes that synchronic AVCs in Sandawe and Hadza indicate that the two languages previously had Verb-Auxiliary order.

2.2 The languages of the study

The Tanzanian Rift Valley Area has been the site of extensive, sustained contact between languages of different linguistic stocks for a very long time indeed. What follows is a brief introduction to the languages of our sample, with languages presented alphabetically within their larger family (if applicable), and arranged from the family with the most numerous members in the sample (Bantu) to the convincingly-argued isolate, Hadza (Figure 1).

Figure 1: Map of the Tanzanian Rift Valley Area (from KMN 2008: 187).

2.2.1 The South Cushitic languages

Forming part of the larger Afroasiatic phylum, the South Cushitic branch numbers 8 spoken (or formerly-spoken) languages, mainly spoken in Tanzania, but with Dahalo (dal) spoken in Kenya. The South Cushitic languages of our sample are: Alagwa (wbj); Burunge (bds); Gorwaa (gow); and Iraqw (irk). Internal classification of South Cushitic is difficult (Kießling and Mous 2003: 2–3): Qwadza (wka) and Aasax (aas) are dormant, with only small amounts of extant lexical data; Ma'a (mhd) is a mixed language (Mous 1994) with the Cushitic component limited to the roots of the 'inner' register only; and the status of Dahalo is unclear. With that said, within the largest branch of South Cushitic (named West Rift in Kießling and Mous 2003), there has been considerable success in representing genetic affiliation (Figure 2).

Alagwa, also known as Uasi, is spoken in and around the Kondoa district by approximately 52,800 people (LOT 2009: 2). Neighbouring languages include Rangi, Datooga, Gogo [gog], and Sandawe. All Alagwa data used in this paper are from Mous (2016).

Burunge is spoken by approximately 23,000 people (LOT 2009: 2) in and around the Kondoa and Chemba districts of Dodoma, central Tanzania. Neighbouring languages include Rangi, Sandawe, Datooga, Gogo, and Maasai [mas]. All Burunge data used in this paper are from Kießling (1994).

Gorwaa, also known as Fiomi, is spoken by approximately 133,000 people in the lowlands surrounding Lake Babati (Harvey 2019a: 139–141), and has a high degree of mutual intelligibility with Iraqw. Contemporary contact with speakers of Iraqw and Mbugwe to the north, and speakers of Rangi and Alagwa to the south and east is strong and frequent. Evidence from barely two generations ago also shows close contact between speakers of Gorwaa and the Datooga varieties (for a more detailed treatment, see Harvey (2019a: 137–138)). Gorwaa data used here come from Harvey (2017).
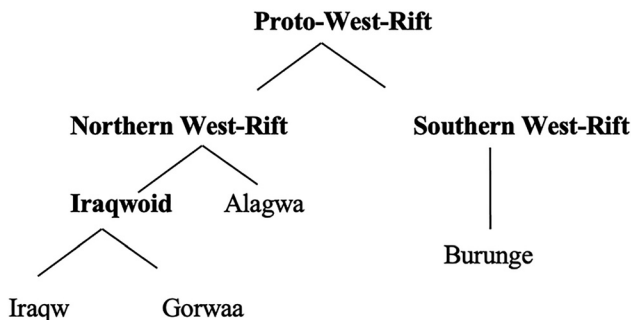

Figure 2: Internal classification of West-Rift (adapted from Kießling and Mous 2003: 2).

Iraqw, also known as Mbulu, is spoken by approximately 602,600 people (LOT 2009: 2), primarily in the highlands of north-central Tanzania. Mutual intelligibility with Gorwaa is high. Neighbouring languages include Maasai, Datooga, Hadza, Gorwaa, Mbugwe, and Nyilamba. All Iraqw data used in this paper are from Mous (1993, 2007). In this paper, Iraqw and Gorwaa data are treated together.

Typologically speaking, the South Cushitic languages Iraqw, Gorwaa, and Alagwa exhibit predominantly SOV word order, with other orderings possible in both Gorwaa and Alagwa for pragmatic effects. Burunge is predominantly SVO, with similar flexibility as Gorwaa and Alagwa. The languages all employ a system of grammatical gender, which is conflated with number in a series of nominal suffixes, triggering agreement on dependents including adjectives and verbs. The languages are primarily head-marking, fusional, and employ a series of verbal extensions which alter the lexical semantics of the verb, and sometimes also its valency.

2.2.2 The Bantu languages

The Bantu languages, part of the larger Niger-Congo phylum, are widely spoken by people throughout central, eastern and southern Africa. The Bantu languages of our sample are Ihanzu [isn], Mbugwe [mgz], Nyaturu [rim], Nyilamba [nim], and Rangi [lag]. All of these languages form part of Guthrie's F30 group (Maho 2009: 45),[2] but when attempting a genetic grouping, the membership of Rangi and Mbugwe (though convincingly shown to be related to each other) is generally contested (c.f. Dunham 2005; Mous 2021; Stegen 2003). Save these two languages, below is a genetic tree placing the rest of the Bantu languages within a subgroup referred to as Takama (Nurse and Philippson 2003). Nodes labelled X represent unnamed predecessor languages of lower branches (Figure 3).

Ihanzu, also known as Isanzu, is spoken by approximately 26,000 people (LOT 2009: 2) in Mkalama district in northern Singida region. Neighboring languages include Nyilamba, Hadza, Datooga, and Sukuma [suk]. All Ihanzu data in this paper are taken from Harvey (2019b). In this paper, the Ihanzu and Nyilamba data are treated together.

Mbugwe (also known as Buwe) is spoken by approximately 37,000 people (LOT 2009: 2) in Manyara region. Neighboring languages include Maasai, Iraqw, and

---

2 The Bantu languages are commonly referred to using a system of so-called Guthrie codes. This is a geographical referential system (rather than a genetic one) which is based on the groupings or 'zones' identified by Guthrie (1967/71) and updated by Maho (2009) which divides the Bantu-speaking area into 15 geographic zones, with languages associated with specific zones. The languages are referred to using a letter which identifies the zone in question and then a two-place digit, with the first number representing the sub-group and the second number representing the individual language.
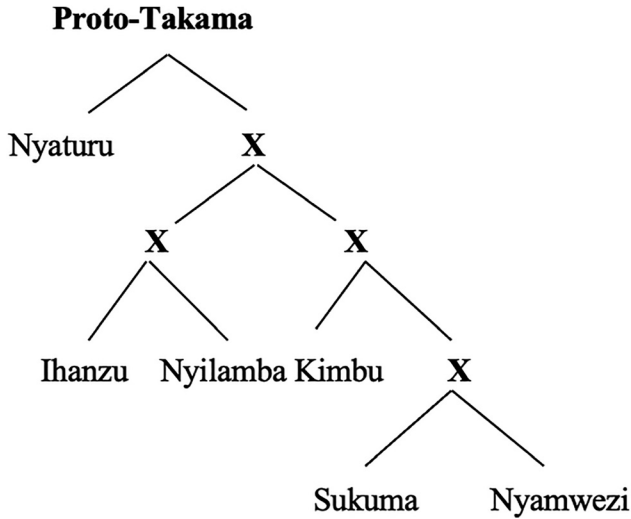
Figure 3: The Takama genetic grouping (adapted from Masele 2001: 274).

Gorwaa. Mbugwe data in this paper are taken from Mous (2000) and Wilhelmsen (2014, 2018).

Rangi (also known as Langi) is spoken by approximately 371,000 people (LOT 2009: 2) in and around Kondoa district of Dodoma region. Two distinct but mutually intelligible varieties – Highland Rangi, and Lowland Rangi – are identified. Neighbouring languages include Alagwa, Gorwaa, Burunge, Sandawe, and Nyaturu. All data in this paper will draw from the Highland variety (Gibson 2012, 2019). In this paper, Rangi and Mbugwe data are treated together.

Nyaturu (also known as Rimi), is spoken by approximately 552,000 people (LOT 2009: 2) across a vast area of central and southern Singida region. Distinct, but mutually intelligible, varieties of Nyaturu include Munyiganyi, Ahi, and Rwana (Masele 2001). Neighbouring languages include Sandawe, Rangi, Nyilamba, Sukuma, Gogo, Nyamwezi, and Kimbu. Olson (1964) describes Rwana as the variety spoken by the most people, and, unless otherwise specified, it is on this variety that the grammar is based. In turn, it is this grammar from which all of our examples are drawn.

Nyilamba (also known as Iramba) is spoken by approximately 386,000 people (LOT 2009: 2) across much of northern Singida region. Ushoola is identified as a distinct but mutually intelligible variety of Nyilamba (Masele 2001). Neighbouring languages include Nyaturu, Ihanzu, Sukuma, and Nyamwezi. Johnson (1923) does not indicate on which variety his grammar notes are based, but it is from this work that all Nyilamba examples will be drawn.

As an aside, Kimbu (kiv) is another Bantu language featured in KMN (2008)'s sample, though at that time, the data on Kimbu were sparse. Unfortunately, that state of affairs remains, and as such, Kimbu is not discussed in the present paper.[3]

The Bantu languages exhibit a broad degree of typological similarity. This is particularly true of the Eastern Bantu languages which have a dominant SVO word order which often allows some flexibility for pragmatic purposes. The languages employ a system of noun classes which behave like grammatical genders and trigger agreement across a range of dependents, including in both the verbal and nominal domain. The languages are primarily head-marking, agglutinative, and make extensive use of a system of verbal suffixes or 'extensions' which perform a range of functions, including in some instances being valency-altering (Figure 4).

2.2.3 The Datooga varieties

Datooga [tcc] is a dialect cluster or group of closely related languages belonging to the Southern Nilotic family (Griscom 2019; Rottland 1982). The Datooga varieties, including Asimjeeg, Barabaiga, Bianjida, and Gisamjanga, are spoken primarily in northern and central Tanzania, and the total number of Datooga speakers is estimated to be 87,800 (Lewis et al. 2013). Mutual intelligibility between varieties varies. The Datooga varieties in our sample are Asimjeeg, and Gisamjanga. Internal classification of the Datooga varieties has not been undertaken since Rottland (1982), whose presentation was meant as a quick sketch.

Asimjeeg, also known as Isimjeeg, is spoken by approximately 3,000 people (Griscom 2019: 5), primarily in four communities: three of which are in the Eyasi Basin, and one far north in the Mara region. Asimjeeg speakers are in regular contact with speakers of Iraqw, Sukuma, and other varieties of Datooga. Asimjeeg speakers often can understand speakers of other varieties of Datooga, but they themselves are not easily understood by speakers of other varieties. Asimjeeg data in this paper comes from Griscom (2019).

Gisamjanga is mutually intelligible with Barabaiga (Mitchell 2021), and together they constitute the largest Datooga group (Schubert et al. 1997: 1). The Gisamjanga and Barabaiga people reside in the Lake Eyasi Basin, the Mbulu Highlands, and in the area surrounding Mt. Hanang. Gismjanga speakers have been in contact with Iraqw speakers for at least a century. Gisamjanga data in this paper comes from KMN (2008), which itself is based on data collected by Paul Berger in the 1930s.

Datooga varieties are generally quite similar typologically, displaying a mixture of agglutinating and fusional morphology, suffixal nominal morphology, and both

---

3 It is our hope, however, that this lack of data will be remedied by Augustino Amos Kagwema's recent (2020) documentary work.

```
                 Southern
                  Nilotic

      Kalenjin              Omotik-
                            Datooga

       Kipsigis
        Keiyo
        Nandi              Datooga        †Omotik
        Terik
        Tugen              Barabaiga
      Kupsabiny             Bajuta
        Sabaot             Bianjida
      Markweeta           Gisamjanga
        Pökoot            Rootigenga
         Akie             Asimjeeg
        Okiek             Buradiga
                            (etc.)
```
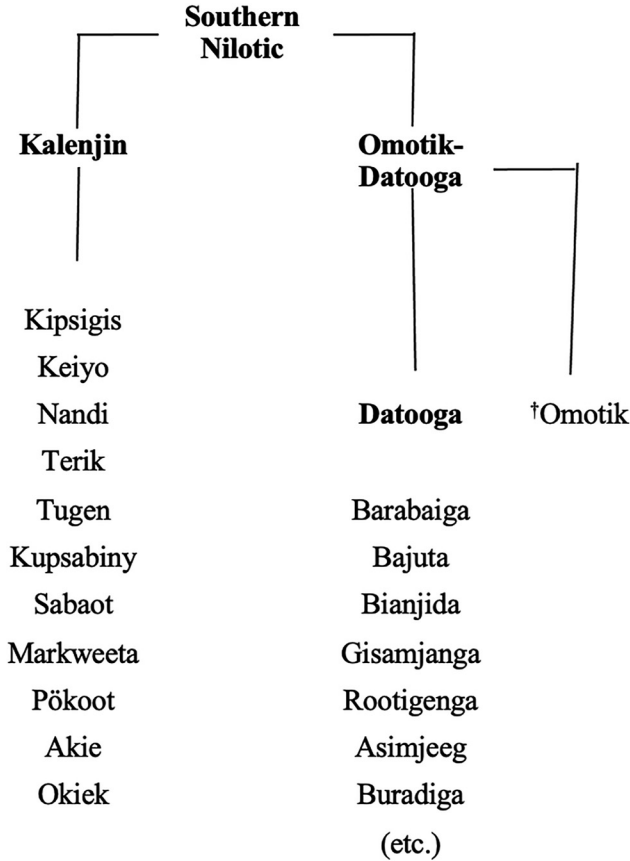
Figure 4: Southern Nilotic family tree (without sub-categorization of Kalenjin).

prefixal and suffixal verbal morphology. Datooga varieties can be said to generally exhibit head-dependent constituent order, whereby prepositions precede nouns, genitive or possessive modifiers follow nouns, and objects typically follow the verb. Datooga varieties have been described as verb-initial, but with flexibility in the constituent order, and Asimjeeg Datooga is described by Griscom (2019) as predominately AVO/SV.

2.2.4 Sandawe

Sandawe [sad] is a language of central Tanzania which is typically classified as either "Khoisan" – i.e. related in some way to other click languages of southern Africa

(specifically the Khoe-Kwadi family) – or unclassified (Sands 1995: 193–4). Sandawe is spoken by approximately 40,000 people and has two mutually intelligible varieties, Eastern Sandawe and Western Sandawe (Eaton 2008: 5). Languages currently in contact with Sandawe include Nyaturu, Rangi, Burunge, and Gogo. Sandawe data used in this paper are from Eaton (2008).

In terms of typological features, Sandawe exhibits a mixture of constituent order patterns. Post-positional and possessive constructions follow dependent-head order, but other nominal modifiers follow head-dependent order (Steeman 2012: 75). Clausal arguments most commonly follow the AOV/SV order in Sandawe, but there is much flexibility (Steeman 2012: 106). Sandawe morphology is primarily suffixing, with a mixture of agglutinating and fusional patterns. Sandawe has two grammatical genders in the singular, which trigger agreement in subject pronominal clitics and nominal modifiers.

2.2.5 Hadza

Hadza [hts] is a language isolate spoken in the Lake Eyasi Basin of northern Tanzania by approximately 1,000 people (Blurton Jones 2016; Edenmyr 2004). Previously thought to be "Khoisan" – i.e. related in some way to other click languages of southern Africa, Sands (1995) establishes Hadza as a language isolate. Hadza is spoken in an area including and bordering Datooga, Iraqw, Ihanzu, Nyilamba, and Sukuma speech communities. Hadza data used in this paper are from Sands (2013a, 2013b, 2013c) and Griscom and Harvey (2020).

Typologically, Hadza shows a mixture of agglutinating and fusional morphology, with primarily suffixing nominal and verbal morphology, but also a small set of prefixes and a single infix. Clause-level and phrase-level constituent order is highly variable but with a tendency for nominal-modifier and auxiliary-verb constituent orders, both of which are associated with but not restricted to VO languages (Dryer 1991). Hadza features two grammatical genders which trigger agreement on nominal modifiers as well as in subject and object verbal indexation (Edenmyr 2004).

3 Preverbal clitic clusters in the Tanzanian Rift Valley: the data

In this section we will explore the preverbal clitic clusters in all twelve of the languages under examination. The languages and the respective qualifying PCCs are

presented in turn, before the possibility and motivation for the range of analyses for the origins of these forms are presented in Section 4.

Though KMN (2008) does not propose a series of criteria by which they identify a PCC, through the structures they include we can identify several properties, the prototypical characteristics of each of which will be listed and exemplified below.

Syntactic independence: prototypically, the preverbal clitic cluster is clearly clitic in nature, in that either material may intervene between the clitic cluster and the verb as in (1), or the clitic cluster itself may occur in more than one syntactic position, as in (2).

(1) *nɪ́ ɪkɪ́i* **njololo** *ɪ́nakʉnkʉa […]*
   nɪ́= ɪ= kɪ́i **njololo** ɪ́-na-kʉnkʉ -a
   REL= SM9= Prst **rooster** SM9-NEG-crow-FV
   'When the rooster had not yet crowed […]' (Nyaturu; Nurse 2000: 523)

(2) a. *kwa**ta** nxaehe baheya migiroda*
      kwa=**ta**   nxae-he   bahe-ya   migiroda/
      COND.AUX=**1.SG.SBJ.ANT** hear-HAB exist-3.M.SG.SBJ.ANT taboo
      'I (often) hear there is a taboo…' (Hadza; Griscom and Harvey 2020 [20200206_BPb_02])

   b. *hakabita **ba** qeqeke atcho*
      haka=bita   **ba**   qe-qeke   atcho
      go=1.PL.INCL.ANT **1.PL.INCL.SBJ.POST** EMPH-cut.3.M.SG.DO skin
      'We will go and really cut the skin' (Hadza; Sands 2013: 269)

Clustering of multiple morphemes: prototypically, the preverbal clitic cluster is (or, at least, can be) composed of the agglutination (as in (3)) or fusion (as in (4)) of multiple morphemes.

(3) *ɪmpíti ʉmʉhʉmba **nɪshánga ázʉmɪkʉwile***
   ɪmpíti   ʉmʉhʉmba   **nʼ=shánga=áza=ʉ-mɪ-kʉw-ile**
   hyaena   boy   **REL= NEG= PST2=SM.3RD.SG-OM.9-hit-PRF**
   'the hyaena that the boy didn't hit' (Ihanzu; Harvey 2019b [20180519b.12])

(4) *aayooríyâ buura **a** leesá aansint naa/asa […]*
   aayi -oo -ríyâ   buura   **i= a=** ∅   leesá   aansint   naa/asa
   mother-F-1SG.POSS   beer   **s.3= O.F=AUX** first   start:3F   brewing
   'my mother first started brewing beer […]' (Alagwa; Maarten Mous p.c. 30.11.2020)

Characteristic semantic domains: prototypically, the preverbal clitic cluster encodes several salient semantic concepts, including subject argument, object argument,

case, tense, clause type, sequentiality, and/or focus. Example (5) shows tense and subject argument marking, and example (6) shows subject argument marking, as well as tense-aspect-mood.

(5) *wun maang'ol **gida**chagsiineeny*
ø-wún   má:ŋòl   **g-ì-dà**-tʃàg-sí:n-é:ɲ
2.sg-come   Mang'ola   **AFF-FUT-1.sg**-send-TERM-2.sg
'Come, I'm going to send you to Mang'ola' (Asimjeeg Datooga; Griscom 2018: [IGS0229_2017- 3-15_#04_031])

(6) *za**ta***
za=**ta**
come=**1.sg.sbj.ant**
'I will come.' (Hadza; Sands 2013: 117)

Distribution: prototypically, the preverbal clitic cluster is obligatory in all finite clauses. Example (7) shows that the finite clause 'the man hit the boy' is grammatical when the preverbal clitic complex is present (a), and ungrammatical when it is omitted (b).

(7) a. *hhawata garma **nguna** taáhh*
hhawata   garma   **ng= u= ∅ =na**   **t**aáhh ~$B~ ~ ~ -a
man   boy   A.3=P.M= AUX =IMPRF   hit ~3~ ~PST~ -FV
'The man hit the boy.' (Gorwaa; Harvey 2017 [20160119f.39])

  b. **hhawata garma taáhh*
hhawata   garma   taáhh ~$B~ ~ ~ -a
man   boy   hit ~3~ ~PST~-FV
Intended: 'The man hit the boy.' (Gorwaa)

Note that not all preverbal clitic clusters in our sample display all of the characteristics described above: these are merely prototypes used to help identify and characterize the structures we wish to focus on. The languages in Section 3 are ordered according to how prototypical their preverbal clitic clusters are, beginning with Iraqw-Gorwaa, whose preverbal clitic complexes are the most prototypical, to Rangi-Mbugwe, whose preverbal clitic complexes are the least prototypical.

3.1 Iraqw-Gorwaa (South Cushitic)

The analyses for Gorwaa (Harvey 2018) and Iraqw (Mous 1993, 2005) differ, but the forms are virtually the same. As such, data for these languages will be treated

together in this paper, with roughly as per Harvey (2018).[4] Every finite clause in Gorwaa and Iraqw contains a preverbal clitic cluster[5] (see (8)).

(8)   *hhawata garma **nguna** taáhh*
       hhawata   garma   **ng**= **u**= ∅ =**na**   taáhh ~$B~ ~ ~ -a
       man   boy   **A.3**=**P.M**=AUX=IMPRF   hit ~3~ ~PST~ -FV
       'The man hit the boy.' (Gorwaa; Harvey 2017 [20160119f.39])

Phonologically, preverbal clitic clusters can bear no tone, which distinguishes them from lexical categories, but, syntactically, these forms are independent, with constituents including patient arguments, oblique arguments, and adverbs able to occur between the preverbal clitic cluster and the verb (see (9)).

(9)   *inós i **hhartá hhawati malé** hanmiis*
       inós   i= ∅   **hhartá**   **hhawata=i**   **malé**   hanmiís ~$B~ -a
       PRO.3SG   S.3= AUX   **stick.LFT** **man=ABL**   **again**   give ~3~ -FV
       'He is giving a stick to the man again' (Iraqw; Mous 2007: 25)

Morphosyntactically, the preverbal clitic cluster is a series of clitics either procliticised or encliticised to a semantically null auxiliary. When the auxiliary has no phonologically-realised argument markers, it is realised as **a**. Schematically, the preverbal clitic cluster selector may be illustrated as seen in Table 1 (where elements within the same column are mutually exclusive of each other).

Gorwaa and Iraqw index all core arguments as proclitics to the auxiliary (see Table 2 below). Morphosyntactic alignment is split, depending on whether the argument is third person, or a speech-act participant (i.e. 1st or 2nd person). For third person arguments, alignment is (superficially) tripartite: the (S)ole argument of an intransitive clause, the (A)gent of a transitive clause, and the (P)atient of a transitive

---

4 The choice of the analysis of Harvey (2018, specifically pp. 137–162), over that of Mous (1993, 2005, see especially 1993: 123–154) is not trivial. One primary difference between Mous and Harvey is what exactly they view the clitics of these languages cliticising to. Where Mous posits a number of forms meaning roughly "to be" to which the various clitics attach, Harvey posits a single, semantically null (and sometimes also phonetically null) auxiliary to which the various clitics attach, and elaborates a few largely morphophonological rules which produce the surface forms. An anonymous reviewer points out that the reasoning behind positing such a null element is sometimes unconvincing (especially in verbal predication), to which we respond that this is entirely valid, especially pending a more thorough, dedicated analysis in a future publication. With that said, we believe that Harvey's (2018) approach handles the complex data in a unified manner which may be more explanatory than Mous' (1993, 2005) approach. If this clarification strikes the reader as rather promissory in tone, we refer them to the publications and page references above, where they can more directly compare each method's individual merits. Note also that Harvey's (2018) analysis is also adapted for both the Alagwa and Burunge data in this paper.
5 Except for the finite imperative, which does not have a preverbal clitic cluster.

Table 1: A schematic representation of the preverbal clitic cluster in Iraqw-Gorwaa.

Mood	Voice	Arguments	Auxiliary	Aspect	Other
Indicative	Active	S	∅	Perfective	Reason
∅=	∅=	A | P		=(g)a	=s
Conditional	Mediopassive			Imperfective	Lative
bar=	t=			=na	=i
Prohibitive				Expectational	Ablative
m=				=n	=wa
Questioning				Consecutive	Instrumental
m=				=re	=r
				Background	
				=wa	

Table 2: A schematic representation of the argument-marking portion of the preverbal clitic cluster in Iraqw-Gorwaa.

	S	A	P		
1			Sg		ti=
			Pl		tindi=
					OR ti= in irk
2	∅=		Sg	M	u=
				F	i=
			Pl		nu=
					OR tundu= in gow
3		ng=	M		u=
	i=	OR	F		a=
		g= in irk	N		i=

are all realised differently. For arguments which are speech-act participants, alignment is accusative: the (S)ole argument of an intransitive clause and the (A)gent of a transitive clause are marked in one way, and the (P)atient of a transitive clause are realised differently.

When a direct object intervenes between the preverbal clitic cluster and the lexical verb, it is no longer marked as an argument within the clitic cluster. This phenomenon is sometimes called encapsulation (Kießling 2007; Whiteley 1958), and fits Mithun's (1984) description of Type III incorporation. As can be seen in (10) below, when the noun **dó**' **Ngawdá**' 'the house of Ngawdá' precedes the clitic cluster, it is marked as a (P)atient argument (**u-**). In a similar sentence, (10), a similar argument

do' oo Qutadu 'the house of Qutadu' occurs between the clitic cluster and the lexical verb. In this case, the encapsulated noun phrase is no longer marked on the clitic cluster, the clitic cluster instead showing marking as if it only had an (S) argument – that is, as if it were an intransitive verb.

(10)   a.   *ina do' oo Qutadu káy*

         i= ∅ =**na**   **do'**   **oo**   **Qutadu -ó**   **káw ~$B~ -i´**

         s.3= AUX =IMPRF   house   ANA.M   Qutadu -L.MO   go ~M~ -PST

         'he went to the house of Qutadu' (Gorwaa; Harvey 2017: [20151125j.76])

   b.   *dó' Ngawdá' nguna káy*

         do'-ó   **Ngawdá'-ó**   **ng**= **u**= ∅ =**na**   **káw~$B~ -í**

         house -L.MO   Ngawdá' -L.MO   A.3= P.M= AUX =PRF   go ~M~ -PST

         'he went to the house of Ngawdá' (Gorwaa; Harvey 2017: [20151202e.105])

Finally, if the object is postverbal in Gorwaa and Iraqw, it is still obligatorily marked in the clitic cluster, as in (11).

(11)   a.   *u imu/umaán dó' hatlá'*

         ∅= **u**= ∅   imu/um-aán-a   do'-ó   hatlá'

         A.P= P.M= AUX   begin-1PL-PRES   house-L.MO   other

         'We are starting another house.' (Gorwaa; Harvey 2017: [20210318a.92])

   b.   **a imu/umaán dó' hatlá'*

         ∅= ∅   **imu/um -aán-a**   **do'-ó**   **hatlá'**

         S.P= AUX   begin -1PL-PRES   house-L.MO   other

         Intended: 'We are starting another house.' (Gorwaa; Harvey 2017: [20210318a.93])

This pattern seems to differentiate Gorwaa-Iraqw from the other South Cushitic languages in the sample.

3.2 Alagwa (South Cushitic)

Finite clauses in Alagwa contain a preverbal clitic cluster (12), though clitic clusters which mark only the arguments may be omitted if the subject is phonologically overt (13) (Mous 2016: 173). As for Iraqw and Gorwaa above, the analysis of Alagwa clitic clusters in this paper is roughly as per Harvey (2018).

(12)   *aayooríyâ buura **a** leesá aansint naa/asa […]*

         aayi -oo -ríyâ   buura   **i**= **a**= ∅   leesá   aansint   naa/asa

         mother -F-1SG.POSS   beer   s.3= o.F= AUX   first   start:3F   brewing

         'my mother first started brewing beer […]' (Alagwa; Maarten Mous p.c. 30.11.2020)

(13)　　*hirúk hirad difafin*
　　　　hirú -k　　　hira -d　　　dif -af -in
　　　　man-M.DEM1　people -DEM4　beat -HAB -IMPF:3M
　　　　'This man beats up people.' (Alagwa; Mous 2016: 133)

Phonologically, and different from Gorwaa and Iraqw, at least one of the enclitics of the Alagwa clitic cluster – general past enclitic **-áa** – is marked as having high tone. As with Gorwaa and Iraqw, these forms are syntactically independent, with constituents including direct and oblique object arguments, and adverbs able to occur between the preverbal clitic cluster and the verb, as in (14) below.

(14)　　*garóo'ín ningi **qaroo diití** bu'ut*
　　　　ga -roó -'ín　　ning= i= ∅　　**qaroo　　diití**　bu'ut
　　　　thing -F -POSS　CSEC= S.3= AUX　**already　here**　be.enough:3F
　　　　'Their case ended here.' (Alagwa; Mous 2016: 222)

Morphosyntactically and similar to Gorwaa and Iraqw, the preverbal clitic cluster is a series of clitics either procliticised or encliticised to a semantically null auxiliary. Unlike Gorwaa and Iraqw, however, this auxiliary is always phonologically null.[6] Schematically, the preverbal clitic cluster may be illustrated as shown in Table 3 (where elements within the same column are mutually exclusive of each other).

　　　Alagwa indexes all core arguments as proclitics to the auxiliary (see Table 4). Morphosyntactic alignment is accusative, allowing a division between (S)ubject and (O)bject.

　　　As with Gorwaa and Iraqw, if the direct object is encapsulated between the clitic complex and the verb, it is unmarked on the clitic complex, and the clitic complex behaves as if it is encoding an intransitive sentence (see (15)).

(15)　　*iyaa too losano /ísin*
　　　　i= ∅ =aa　　too　losano　/ís -in
　　　　S.3= AUX =PST　just　initiation　do -IMPF:3M
　　　　'He only organised initiation.' (Alagwa; Mous 2016: 222)

Alagwa has a freer word order than Iraqw: typically, new information is introduced following the lexical verb, and given information is introduced before the clitic complex (Mous 2016: 222). A postverbal object is given in (16) below. If the object is postverbal in Alagwa, it is not marked in the clitic complex.

---

6 One important piece of evidence for this can be seen in the copular constructions of Alagwa (e.g. the examples (9.1) in Mous 2016: 220).

Table 3: A schematic representation of the preverbal clitic cluster in Alagwa.

Mood	Clause type			Arguments	Auxiliary	Tense	Other
Indicative Ø=	Beneficient **s=**	Subordinate **k=**	Impersonal subject **k=**	S | O	Ø	General past **=áa**	Ablative **=aa**
Optative l= OR **n=**						Perfect **=n(V)**	Instrumental **=ra**
Consecutive **n=**						Predicative focus **=n**	
Ventive **n=**							
Immediate **n=**							

Table 4: A schematic representation of the argument-marking portion of the preverbal clitic cluster in Alagwa (Adapted from Mous 2016: 174).

	S		**O**
1	**a=**	M	**i=**
		Pl	**kunu=**
2		M	**oo=**
		F	**i=**
		Pl	**kunu=**
3	**i=**	M	**oo=**
		F	**a=**
		Pl	**i=**

(16)   *Ama Irimi **iyaa** gu/umint hiru wak*
       Ama   Irimi   **i=** ∅ =**aa**   gu/ -ut   hiru   wak
       Ama   Irimi (ogre) s.3= AUX =PST swallow -3F man   one
       'Ama Irimi swallowed a man.' (Alagwa; Mous 2016: 133)

3.3 Burunge (South Cushitic)

Finite clauses in Burunge contain a preverbal clitic cluster (17). However, unlike the other South Cushitic languages, the most common constituent-ordering in Burunge is Subject-Verb-Object, and in this configuration, marking the object in the preverbal clitic complex is optional (compare (18a) in which the object is unmarked, and (18b), in which the object is optionally marked). Analysis of Burunge clitic complexes in this paper is roughly as per Harvey (2018).

(17)   *Laa'ay puncacee wa/ᵃ **higigi** slay*
       Laa'ay   puncacee   wa/ᵃ   **hi= gi=** ∅ =**gi**   slay
       Laa'ay   sheep(pl.)   many   s.3= o.PL= AUX =SEQ   get.3SG.M.PRF
       'Laa'ay got many sheep' (Burunge; Kießling 1994: 206)

(18)   a.   *Laa'ay **higi** slay puncacee wa/ᵃ*
            Laa'ay   **hi=** ∅ =**gi**   slay   puncacee   wa/ᵃ
            Laa'ay   s.3= AUX =SEQ   get.3SG.M.PRF   sheep.(pl.)   many
            'Laa'ay got many sheep' (Burunge; Kießling 1994: 206)
       b.   *Laa'ay **higigi** slay puncacee wa/ᵃ*
            Laa'ay   **hi= gi=** ∅ =**gi**   slay   puncacee   wa/ᵃ
            Laa'ay   s.3= o.PL= AUX =SEQ   get.3SG.M.PRF   sheep.(pl.)   many
            'Laa'ay got many sheep' (Burunge; Kießling 1994: 206)

Morphosyntactically and similar to Gorwaa and Iraqw, the preverbal clitic cluster is a series of clitics either procliticised or encliticised to a semantically-null auxiliary. When the auxiliary has no phonologically-realised argument markers, it is realised as *i*.[7] Schematically, the preverbal clitic cluster may be illustrated as shown in Table 5.[8]

As Alagwa, Burunge indexes all core arguments as proclitics to the auxiliary (see Table 6). Morphosyntactic alignment is accusative, allowing a division between (S)ubject and (O)bject.

As with the other South Cushitic languages, if the direct object is encapsulated between the clitic complex and the verb, it is unmarked on the clitic complex, and the clitic complex behaves as if it is encoding an intransitive sentence (19).

(19)   *Laa'ay higi puncacee slay*
         Laa'ay   hi= ∅ **=gi   puncacee   slay**
         Laa'ay   s.3= AUX =SEQ   sheep   get.3SG.M.PRF
         'Laa'ay got sheep' (Burunge; Kießling 1994: 206)

3.4 Nyaturu (Bantu)

Preverbal clitics do not occur in every clause in Nyaturu (20), but are used to express certain tenses-aspects (sequential, persistive, near and far past, and near and far future) (21), as well as for subordinate clauses (22).

(20)   *ng'ombe atamʊe ígwe*
         ng'ombe   ʊ-a-tamʊ-íe   ígwe
         c9.cow   SM9-PST-split-PERF   c5.boulder
         'a cow has split a boulder' (Nyaturu; Olson 1964: 157)

(21)   *njotá **iqaá** ʊgwa*
         njotá   **i= qaá**   ʊ-gwa
         c10.stars   SM10= SEQ   SM10- fall
         '[and] the stars will fall' (Nyaturu: Olson 1964: 212)

---

7 Evidence for this form comes in the copular constructions of Burunge, in which the "bare" auxiliary emerges as **yaa (i =aa** (Aux =Fut1), **yóo (i =óo),** and **yoo (i =oo)**.
8 Unlike the other tables given for South Cushitic, it is not certain whether material which occupies a single column (unless clearly semantically incompatible) in this table is mutually exclusive or not. For example, Roland Kießling (p.c. 18.01.2021) indicates that he has no counterexamples to rule out the co-occurrence of the subject focus proclitic *na=* and the benefactive *sa*.

Table 5: A schematic representation of the preverbal clitic cluster in Burunge.

Mood	Clause type	Arguments	Auxiliary		Tense and aktionsart	Direction of action
Indicative Ø=	Subject focus **na=**	S | O	Ø	Object focus **=ni** Comitative/instrumental **=ri**	Present **=Ø**	Ventive **=ni**
						Sequential **=gi**
Conditional **bara=**	Object relative **ga=**				Preterite **=áa**	Reflexive **=ti**
Optative **la=**	Indefinite subject **da**				Completive **=ng**	Separative **=ti**
	Benefactive **sa**				Future 1 **=aa**	
					Future 2 **=maa**	
					Habitual **=óo**	
					Prospective **=oo**	

Table 6: A schematic representation of the argument-marking portion of the pre-verbal clitic cluster in Alagwa (Adapted from Mous 2016: 174).

	S			**O**
1	**ha=**	Sg		**ni=**
		Pl		**ndi=**
2		Sg	M	**gu=**
			F	**gi=**
		Pl		**ngu=**
3	**hi= OR ∅=**	M		**gu=**
		F		**ga=**
		N		**gi=**

(22) *mʊɾimányíé **nɨ́**mʊʉɾɪromba*
mʊ-ɾi-mány-íé   **nɨ́**=mʊ-qʉ́-rɪ-romb-a
SM2-OM5-know-PERF   REL=SM2-PROG-OM5-ask-FV
'You (pl.) don't know what you are asking for' (Nyaturu; Olson 1964: 136)

When these tense-aspects and clause types are expressed together, these preverbal clitics occur in what KMN (2008) identify as the clitic cluster for this language. This exemplified in (23) below, where subordination and far past are marked in the clitic cluster.

(23) **nɨ́ náa akɨ́ɨ** *ʉqʉrighiRya*
nɨ́ = **náa**   a= **kɨ́ɨ**   ʉ́-qʉ-righiRy-a
REL=   FPST   SM3RD= PRST   SM3RD-PROG-speak-FV
'while she was still speaking […]' (Nyaturu: Nurse 2000: 523)

Phonologically, these preverbal clitic complexes can bear tone. Syntactically, these forms are independent, with constituents such as the subject of an intransitive verb (both unergative (24) and unaccusative (25) able to intervene between what has been described as the preverbal clitic cluster and the verb.

(24) *nɨ́ ɨkɨ́ɨ **njololo** ɨnakʉnkʉa*
nɨ́=   ɨ= kɨ́ɨ   **njololo**   ɨ- na- kʉnkʉ -a
REL=   SM9= Prst   **rooster**   SM9-NEG-crow-FV
"When the rooster had not yet crowed […]" (Nyaturu: Nurse 2000: 523)

(25) *nɨ́ náa **Ntandʉ** wakuya*
nɨ́=   náa   **Ntandʉ**   ʉ-a-kuy-a
REL=   FPST   **Ntandʉ**   SM1-FPST-die-FV
'When Ntandʉ died' (Nyaturu: Olson 1964: 202)

The Nyaturu preverbal clitic cluster can be analysed as analogous to a series of additional verbal 'positions' which occur before the initial part of the verb,

traditionally construed. The structure of the verbal template, following Güldemann (1999: 546) is shown in (26) below.

(26)

Pre-initial	**Initial**	**Post-Initial**	**Pre-Radical**	**Radical**
TAM/Polarity	SM	TAM/polarity	OM	verb root

Pre-Final	**Final**	**Post-Final**
derivation	TAM	clause-type
TAM		polarity

Under the system found in Nyaturu the first element in the verb phrase is occupied by the relative marker **ní**, which can also be considered as a proclitic (cf (3) vs. (6) above).[9] The tense-indicating auxiliaries (e.g. the far past **náa**) may occur next, in a position which is most probably the pre-initial position of (26) immediately above. These so-called tense-indicating auxiliaries are used to encode sequential senses, connecting one clause to the next (cf. similar systems that are found in South Cushitic in example (14) above). In terms of linear ordering, the auxiliaries which appear after those encoding tense indicate aspect or clause-combining. And it is only after this that auxiliaries optionally marking the subject of the verb occur. But notably they use a different inflectional paradigm than that employed for subject marking on the main verb, reflecting a widely documented morphosyntactic difference between dependent and independent clauses in Bantu languages (Güldemann 1999). Schematically, the preverbal clitic complex can be illustrated as shown in Table 7.

Table 7: A schematic representation of the preverbal clitic cluster in Nyaturu.

Clause type	Tense	Aspect/clause-combining			
		(Optional) Subject argument			
		Sg.	Pl.		
Main clause ∅=	Far past **náa**	1st	**N-**	**qʊ-**	Sequential **qàá**
Subordinate **ní=**	Near past **ájà**	2nd	**ʊ-**	**mʊ-**	Persistive **kɨ̀ɨ̀**
	Near future **nàa**	3rd	**a-**	**vɨ-**	
	Far future **ìkwí**	*Classes 1–18 inflect as per the main verb			

---

9 Of **ní**, an anonymous reviewer observes that this may not be a verbal clitic as such, but rather a phrasal relative complemetiser. This view probably also applies to the analogous forms in Ihanzu and Nyilamba. With this said, all of the forms occur preverbally and seem to behave enough like clitics to merit their inclusion here.

Note that, of the (tense) auxiliaries, a future auxiliary may occur simultaneously with a past auxiliary. The result is an event which happened in the past, but subsequent to something else (27). This does not seem to be the case for slot 3 (aspect/clause-combining) auxiliaries.

(27)   *nɪ́ **náa ikwɪ́** waʋfenja*
        nɪ́=   **náa**   **ikwɪ́**   ʋ-a-ʋfenj-a
        REL=   FPST   FFUT   SM.3RDSG-FPST-want-FV
        'when he afterwards wanted' (Nyaturu; Olson 1964: 204)

The slot 2 (tense) auxiliaries may also occur without a main verb in copular constructions (28).[10] The same pattern has not been described for the aspect/clause-combining auxiliaries.

(28)   *qʋsóko **náa** mʋkufɪ*
        qʋsóko   **náa**   mʋkufɪ
        because   FPST   c1.short
        'because he was a short man" (Nyaturu; Olson 1964: 202)

3.5 Datooga (Southern Nilotic)

Verbs in all of the Datooga varieties include both preverbal and post-verbal morphology (Kießling 2007; Rottland 1982). Griscom (2019) identifies two sets of preverbal morphology in Asimjeeg Datooga associated with two distinct categories of verbal constructions: *simplex-stem* and *complex-stem* verbal constructions. Simplex-stem constructions maximally feature three prefixes (see Table 8 while complex-stem constructions may feature up to five (see Table 9). The first four slots of the corresponding preverbal morphology of complex-stem constructions in Gisamjanga Datooga are identified together by KMN (2008) as a preverbal clitic cluster. These include the 1) conditional, 2) polarity, 3) future tense, and 4) aspect slots. The morphemes in these slots are represented here as prefixes here (see Section 4.2).

The simplex- and complex-stem constructions feature slightly different subject indexation paradigms and are utilized in two mutually exclusive sets of tense-aspect-polarity constructions (see Table 10). Within each of the complex-stem constructions, at least one morpheme from the first four preverbal morphology slots must be present.

---

10 An anonymous reviewer points out that this pattern of copular use for these auxiliaries may point to a path of grammaticalisation, where the occurrence of forms like **náa** as copulae was the basis for their use as auxiliaries. This would be commensurate with patterns in other Bantu languages, such as the Swahili verb **kuwa** 'to be', which can be used as a copula (**alikuwa tajiri** 'he was rich'), and as an auxiliary (*alikuwa ameamka* 'he had woken up').

Table 8: Simplex-stem preverbal morphology in Asimjeeg Datooga.

(Conditional)	(Affirmative/negative/temporal)	Subject indexation	Verb root
ì:(j)-	**g- ~ q-m-/am-**	–	–[a]

[a]Morphological slots whose contents are marked with the symbol "–" consist of either an open lexical class (i.e., verb root) or a morphological paradigm (i.e., subject indexation).

Table 9: Complex-stem preverbal morphology in Asimjeeg Datooga.

1 (Conditional)	2 (Affirmative/ negative)	3 (Future)	4 (Aspect)	Subject indexation (Affirmative)	Verb root
ì:(j)-	**g- ~ q-/m-**	**i(dʒ)(a)-**	**ad-** persistive **gòl-** affirmative priority	–	–

Table 10: Simplex- and complex-stem constructions in Asimjeeg Datooga (Griscom 2019).

Simplex-stem constructions	Complex-stem constructions
Non-future tense	Future tense
Affirmative perfect tense-aspect	Persistive aspect
	Affirmative priority aspect
	Negative perfect aspect

The difference between the simplex- and complex-stem subject indexation paradigms are exemplified in (29) and (30). In (29) the verb root **ŋɛ́:d** 'wake, start' occurs in a non-future tense construction with simplex-stem morphology, including the affirmative prefix **q-** and the first person subject indexation marker **á:-**. In (30), by contrast, the verb root **tʃàg** 'send', occurs in a future tense construction with complex-stem morphology, including the affirmative prefix **g-**, the future prefix **ì-**, and the first person subject indexation marker **dà-**.

(29)   *aniin qay **qaa**daedaew mang'ol*
        àní:n   qáj   **q-á**:-ŋɛ́:d-ɛ̀:w   má:ŋòl
        1.SG.PRO   old.times   AFF-**1.SG**-wake-LOC   Mang'ola
        'In the old times. I began in Mang'ola.' (Asimjeeg Datooga; Griscom 2018: [IGS0229_2015-12- 21_GG_01_003])

(30)   *wun mang'ol **gida**chagsiineeny*
        ø-wún   má:ŋòl   **g-ì-dà**-tʃàg-sí:n-é:ɲ
        2.SG-come   Mang'ola   AFF-FUT-**1.SG**-send-TERM-2.SG
        'Come, I'm going to send you to Mang'ola' (Asimjeeg Datooga; Griscom 2018: [IGS0229_2017- 3-15_#04_031])

KMN (2008) analyze the future tense morpheme **i(dʒ)(a)-** as a clitic based on examples of two different constructions occurring in the Berger corpus (recorded by Paul Berger in 1935/1936). The first of these is analyzed as a future relative construction (see (31). For Asimjeeg Datooga (see (32), Griscom (2019) offers an alternative analysis as a distinct but etymologically related morpheme due to semantic similarities but morphosyntactic differences, such as the lack of a polarity marker or subject indexation on the future relative, which set the verbal future morpheme and the future relative apart.

(31)   *qwayap hilooga qoohaat heedajaa shinyada gwalla nooga*
       qwàjâp híló︎ogà   qòohâat   héedá jàa   ʃíɳádà gwállà
       s3.erect cattle.enclosure s3SG:INCREASE place FUT.REL evening s3.sleep.at
       nòoga
       goats
       'They built a cattle enclosure to increase the room for the goats and sheep to sleep at night' (Gisamjanga Datooga; KMN 2008: 203)

(32)   *ginuny sinaad eed ja qwalaap beeg*
       g-ì-ø-nùɲ   ø-sìn-áːd   éː-d   dʒá   q-ʷá-láːp béːː-g
       AFF-FUT-2.SG-come 2.SG-do-AM.ITV place-SS.SG FUT.REL AFF-3-pass water-SS.PL
       'You come to prepare a place to pass water through…' (Asimjeeg Datooga; Griscom 2019: [IGS0229_2017-3-2 #1_128], glossing modified)

The second construction KMN (2008) cites in support of the analysis of the future construction as a clitic is also a construction found in the Berger corpus, and one in which the future is separated from the lexical verb by the subject argument (see (33)). Unlike the relative construction, this separated future construction has not been found in any other Datooga data.

(33)   *akaja gaba siisi guursa oorjeedaanyi*
       ák-àjà   gábá   síisí   gùurs-á òorjéedàa-ɲi
       SEQ:AFF-FUT every   person   call:APPL-3son-POSS.3SG
       'then everyone will call his son' (Gisamjanga Datooga; KMN 2008: 203)

3.6 Ihanzu-Nyilamba (Bantu)

Ihanzu and Nyilamba show some significant differences (for lexical differences, see Masele 2001), but in terms of the preverbal clitic cluster, we feel the forms are sufficiently similar such that the data from these two languages may be treated together here. As with Nyaturu, preverbal clitics are not obligatory elements of every

clause in Ihanzu or Nylamba (see the Ihanzu example (34)), but are used to express certain tenses (35), negation (36), as well as for subordinate clauses (37).

(34)   *mʋhʋmba ʋmɩkʋwile ɩmpíti*
         mʋhʋmba   ʋ-mɩ kʋw -ile   ɩmpíti
         boy   SM1-OM.9-hit-PRF   hyaena
         'The boy hit the hyaena.' (Ihanzu; Harvey 2019b: [20180531m.1])

(35)   *ʋmʋhʋmba **álɩ**mɩkʋwile ɩmpíti*
         ʋmʋhʋmba   **álɩ**= ʋ-mɩ-kʋw-ile   ɩmpíti
         boy   PST3= SM1-OM.9-hit-PRF   hyaena
         'The boy hit the hyaena.' (Ihanzu; Harvey 2019b: [20180521f.1])

(36)   ***sika**nitendile*
         sika=ni-tend-ile
         NEG=SM.1SG-do-PRF
         'I didn't do (it).' (Nyilamba; Johnson 1923: 180)

(37)   *sime izo **na**ntakile*
         sime   izo   **na**=n-tak-ile
         9.knife 9.DEM2   REL=SM.1SG-want-PRF
         'That knife which I want.' (Nyilamba; Johnson 1923: 182)

When these notions are expressed together, we may posit the resulting string as the preverbal clitic complex in these languages (38).

(38)   *ɩmpíti ʋmʋhʋmba **nɩshánga áz**ʋmɩkʋwile*
         ɩmpíti   ʋmʋhʋmba   **n´= shánga**= **áza**=ʋ-mɩ-kʋw-ile
         hyaena   boy   REL= NEG=   PST2= SM.3RD.SG-OM.9-hit-PRF
         'the hyaena that the boy didn't hit' (Ihanzu; Harvey 2019b: [20180519b.12])

Phonologically, these preverbal clitic complexes can bear tone. Syntactically, and unlike Nyaturu, material cannot intervene between the clitic complexes and the lexical verb (compare (39)). With that said, the preverbal clitics in Ihanzu do show some degree of flexibility in terms of their relative ordering (compare examples (a) and (b) in (40)), and are therefore not as tightly integrated into the verb as, say, the subject prefixes.

(39)   a.   *ʋmʋgala ntulí názʋkʋle ʋse kalóngola kuBukúndi*
            ʋmʋgala.ntulí   n´=áza=ʋ-kʋ-ile   ʋse   kʋ´-a-longol-a
            alcoholic   REL=PST2=SM3-die-PRF   PRO.1PL   SM.1PL-PST1-depart-FV
            kuBukúndi
            to.Bukundi
            'When the drunk died, we went to Bukundi' (Ihanzu; Harvey 2019b:
            [20201221a.3])

Table 11: A schematic representation of the argument-marking portion of the preverbal clitic cluster in Ihanzu-Nyilamba.

Clause type	Negation	Tense
Main Clause ∅**=**	Positive ∅**=**	Past 2 **áza=**
Subordinate **na=**	Negative **sika=**	Past 3 **álɪ=**
OR **nʹ=** in isn	OR **singa=/shanga=** in isn	

      b.   *náza ʊmʊgala ntulí ʊkʊle ʊse kalóngola kuBukúndi*

           nʹ= áza   ʊmʊgala.ntulí   ʊ- kʊ -ile   ʊse   kʊʹ-a-longol-a

           REL= PST2   alcoholic   SM3- die -PRF   PRO.1PL SM.1PL-PST1- depart-FV

           kuBukúndi

           to.Bukundi

           Intended meaning: 'When the drunk died, we went to Bukundi' (Ihanzu; Harvey 2019b [20201221a.4])

(40) a.   *ɪmpíti ʊmʊhʊmba nɪshánga ázʊmɪkʊwile*

           ɪmpíti   ʊmʊhʊmba   nɪ= shánga= áza= ʊ- mɪ-kʊw -ile

           c9.hyaena   c1.boy   REL= NEG= PST2= SM1- OM9 hit -PRF

           'the hyaena that the boy didn't hit' (Ihanzu; Harvey 2019b: [20180519b.12])

      b.   *ɪmpíti náza sínga nɪmíkʊwile*

           ɪmpíti   n=áza=   sínga= nʹ-mí- kʊw -ile

           c9.hyaena   REL= PST2=   NEG= SM.1ST- OM.9- hit -PRF

           'the hyaena that I didn't hit' (Ihanzu; Harvey 2019b: [20180519b.4])

Morphosyntactically, and within the Bantu convention, the preverbal clitic cluster in Ihanzu and Nyilamba can be analysed as a series of additional verbal 'slots' which occur before the initial part of the verb. Slot 1 is occupied by the relative marker, which is a clear example of a proclitic. Slot 2 is occupied by a marker of negation. Slot 3 is occupied by a tense-marking auxiliary. Schematically, the preverbal clitic complex may be illustrated as shown in Table 11.

3.7 Sandawe (Khoe-Kwadi or isolate)

In Sandawe, the person, gender, and number (PGN) of the subject is coded in realis clauses through verbal enclitics and/or preverbal enclitics that attach to other clause constituents (Eaton 2008). In realis negative and irrealis clauses, there are additional paradigms of verbal suffixes which are distinct from those used in the PGN clitics. In

Table 12: Sandawe person-gender-number (PGN) forms.

	Realis PGN clitics	Realis negative PGN suffixes (high toned)	Irrealis PGN suffixes (low toned)
1.SG	**=sì̧**	**-sé**	**-sì̧**
2.SG	**=ì**	**-pó**	**-pò**
3.M.SG	**=à**	**-é:**	**-ø̀**
3.F.SG	**=sà**	**-sú**	**-sų̀**
1.PL	**=ò**	**-sú̧:**	**-sų̃:, -sà**
2.PL	**=è**	**-sĩ́:**	**-sĩ̀:**
3.PL	**=àʔ**	**-só**	**-sò**

the case of the realis negative clauses, the realis PGN clitic may be present elsewhere in the same clause as a realis negative PGN verbal suffix.[11] The three PGN clitic and suffix paradigms of Sandawe are listed in Table 12.

The PGN clitics can attach to non-subject clause constituents such as objects and adverbs in addition to or rather than the verb, as seen in (41) and (42). The choice of which constituent is marked with the clitic depends on the information structure of the clause (Eaton 2008: 127).

(41) *pà: ⁿ|ʷæ̃̀: kútú:mbî mé:â síé: kòŋkòʔsé: ‖'ò‖'â̂:ts'à: tɬ'àpʰè*
pà: ⁿ|ʷæ̃̀:-:̃   kútú:mbì   mé:=**à**   sí-é-:̃
NC(3M.SG.) elephant-sp. tree.trunk big=**3.M.SG** take-3.M.SG.OBJ-&
kòŋkòʔsé-é-:̃   ‖'ò‖'á-:̃-ts'ì̧=**à**   tɬ'àpʰé
raise-3M.SG.OBJ.-& baboon-sp.-at=**3.M.SG** hit
'Then Elephant took a big tree trunk and raised it up to hit Baboon.'
(Sandawe; Eaton 2008: 127)

(42) *hèwéʔgèsì̧ téɬâsì̧ tʃí kìmæ̃̀:sì̧ ɬá:sì̧ |àní ts'è:ò-nàsì̧ pè:*
hèwéʔgê=**sì̧** téɬè-sì̧   tʃí   kímæ̃́:-:̃=**sì̧**
and.so-=**1.SG** completely-1.SG [1.SG poisonous.arrow]_GEN -SP.=**1.SG**
ɬá:=**sì̧** |àní ts'é:ò-nà=**sì̧**   pě:
well=**1.SG** [bow string]_GEN -to=**1.SG** put
'And so I put my poisonous arrow completely well on the bow string.'
(Sandawe; Eaton 2008: 128)

---

11 With regards to obligatoriness, Helen Eaton (p.c 24.05.2021) notes that the PGN clitic is obligatory in affirmative realis clauses when there is no Subject Focus marker. Interestingly, the PGN clitics and Subject Focus marker look like they might historically have been the same set of independent subject pronouns which then got grammaticalized in two different ways.

The subject NP or pronominal in a realis clause is optionally marked with a subject focus (SF) marker **-á**ː. Generally, a verb without a PGN clitic cannot precede the first PGN clitic or subject-focus (SF) marker of a clause, and a verb with a PGN clitic cannot be preceded by another PGN clitic or SF marker in the same clause (Eaton 2008: 128). In (43), for example, the SF marker precedes a verb without a PC, and in (44) the verb 'go' occurs with a PC prior to all other instances of PCs.

(43) ⁿ|î̀ː tʃʰí**ā**ː ʔíé̋ː kópókòpò
      ⁿ|î̀ː   tʃʰíà=**á**ː   ʔíé -̋ː   kópókópò
      body   all=sꜰ   stay-&   shake
      'the whole body was shaking.' (Sandawe; Eaton 2008: 129)

(44) hík'**ị̀sí̋**ː   mìndàtà**sị̀** |ʷ**ằ**ː**sị̀**̤   ⁿ|è̀ʔw**ằ**ː
      hík'**ị̀**=**sị̀**-̋ː   mìndà-tò=**sị̀** |ʷ**ằ**ː=**sị̀**̤   ⁿ|è̀ː-w**á**ː-̋ː
      go=**1sɢ.ᴘᴄ**-&   field-to=**1sɢ.ᴘᴄ**   millet=**1sɢ.ᴘᴄ**   cut-3ɪ.ᴘʟ.ᴏʙᴊ-&
      'I go and cut millet in the field and' (Sandawe; Eaton 2008: 175)

Furthermore, it is possible that the SF marker and the PGN clitic paradigm may have developed from the same diachronic source, as there are formal similarities between the SF marker and the 3rd person PGN clitics, they share semantic and pragmatic properties, and their distribution is mutually exclusive (Eaton personal communication).

3.8 Hadza (isolate)

In all non-imperative verbal-predicate clauses in Hadza, the person, gender, and number (PGN) of the subject argument is obligatorily coded together with the tense and aspect of the clause. These PGN-TAM morphs occur in four sets of fusional paradigms, identified by Miller et al. (2016) as anterior (non-past or recent past tense), posterior (past or remote past tense), potential (some certainty and/or non-past tense), and veridical (less certainty and/or counterfactual). The segmental forms of one of the PGN paradigms, the anterior, is presented in Table 13. Other PGN paradigms are presented in Section 4.3 below.

Each paradigm may occur either as verbal enclitics as in (45), preverbal enclitics that attach to an auxiliary or adverb as in (46), or as a preverbal syntactically independent constituent as in (47). Enclitic forms exhibit varying levels of phonological dependence and reduction, such as vowel-copying (e.g. **haka** + **ephee** → **haka:phee**) and elision (e.g. **kwa** + **heso** → **kweso**).

Table 13: Hadza person-gender-number (PGN) paradigm for anterior tense.

	Singular	Plural
1st person	**=ta**	**=ota** (EXCL), **=bita** (INCL)
2nd person	**=(t)ita**	**=ete꞉ta** (M), **=(i)ti꞉ta** ~ **=te꞉ta** (F)
3rd person	**=ea** (M), **=ako** (F)	**=ephee** (M), **=iphii** (F)

(45)　　*zata*
　　　　za=**ta**
　　　　come=**1.SG.SBJ.ANT**
　　　　'I will come.' (Hadza; Sands 2013a: 117)

(46)　　*kwata nxae: baheya migiroda*
　　　　kwa=**ta**　　　　　　nxae-e　　　　bahe-ya　　　　　　migiroda
　　　　COND.AUX=**1.SG.SBJ.ANT**　hear-3.M.SG.OBJ　exist-3.M.SG.SUBJ.ANT　taboo
　　　　'I (often) hear there is a taboo …' (Hadza; Griscom and Harvey 2020
　　　　[20200206_BPb_02])

(47)　　*hakabita ba qeqeke atcho*
　　　　haka=bita　　　　　**ba**　　　　　　qe-qeke　　　　　atcho
　　　　go=1.PL.INCL.ANT　**1.PL.INCL.SBJ.POST**　EMPH-cut.3.M.SG.DO　skin
　　　　'We will go and really cut the skin' (Hadza; Sands 2013c: 269)

Sands (2013c: 267) argues that the clitics most commonly attach to an auxiliary (46), and somewhat less frequently attach to the end of the verb (45). The clitics infrequently attach to adverbials or occur as independent constituents. Two clauses with different clitic patterns may be linked together, as in (47). Object indexation is always coded directly as a suffix on the verb, regardless of where or how subject indexation is coded. The Hadza auxiliaries, all of which may occur with PGN clitics, are listed in Table 14, as identified by Miller et al. (2016). The subjunctive occurs with its own paradigm, and for some auxiliaries there are one or more paradigms that have additional habitual forms.

Auxiliaries almost always precede the lexical verb of the same clause, and often directly precede it. Other syntactic units may occur between the auxiliary and the verb, however. In (48) the adverb **kenena** 'early' occurs between the auxiliary and the verb, in (49) the negative auxiliary **akhwe** occurs between the auxiliary and the verb, and in (50) the subject NP **zzutchibisa sanzako** 'the North Wind' occurs between the auxiliary and the verb.

Table 14: Hadza auxiliaries.

Form	Function	PGN-TAM paradigms
(h)a	Sequential	Posterior, veridical
akhwa	Negative	Anterior, posterior, potential, veridical
i	Subjunctive	Subjunctive
ka	Sequential, possible contrastive	Anterior, posterior, veridical
kwa	Relative, conditional	Anterior, posterior, potential, veridical
ya ~ ia	Sequential	Anterior, posterior, veridical

(48)   *kaka kenena era zzoko*

    ka=ka   **kenena**   era   zzoko

    SEQ.AUX=3.M.SBJ.POST   **early**   build   fire

    '… and he had already built his fire.' (Hadza; Bala 1998: 26)

(49)   *kwakwa akhwe samiya paka a hamaisho*

    kwa=kwa   **akhwe**   sam-iya paka a   hamaisho

    COND.AUX=3.F.SBJ.POST   **NEG.AUX** eat-PASS   until even now

    '… she had not eaten up to then.' (Hadza; Bala 1998: 20)

(50)   *beena kitcha zzutchibisa sanzako thaya phoyatcha*

    beena   k=itcha   **zzutchi-bi-sa**   **sanza-ko**

    then   SEQ.AUX=3.F.PL.SBJ.POST   **wind-M.PL-3.F.SG.POSS**   **north-F.SG**

    tha-ya   phoya-tcha

    leave-PASS   blow.wind=3.M.PL.OBJ

    'And then the North Wind stopped blowing.' (Hadza; Griscom and Harvey 2020: [20200306_12])

There is some evidence indicating possible historical origins of the PGN-TAM clitics and auxiliaries. The PGN-TAM clitics can often be subdivided into distinct subject PGN and TAM components, with the PGN component preceding the TAM component in 1st and 2nd person, and the PGN component following the TAM component in 3rd person. This indicates the possibility that Hadza may have previously had Verb-Auxiliary word order, with the TAM morphs coming from verbal auxiliaries and two distinct post-verbal word order patterns for 1st/2nd person and 3rd person subjects. The origins of the synchronic auxiliary system are less clear, but there is evidence of possible connections to semantically and formally similar copula and conjunction constructions.

3.9 Rangi-Mbugwe (Bantu)

Despite their present-day geographical separation, Rangi and Mbugwe are considered sufficiently similar in their clitic clusters for the data from these two languages to be treated together in this paper. For discussion of the preverbal clitics in Rangi-Mbugwe, the relevant construction has been described as a complex auxiliary-based construction. These are used to encode a variety of TAM distinctions, with the specific tense-aspect combinations also showing variation between the two languages. Rangi and Mbugwe were not included in the original study of KMN (2008). We include them here due to the further work that has been done on the languages in the intervening years, as well as to provide a point of reference and comparison for the other Bantu languages included in the current paper. Furthermore, since one of the goals of the current paper is to consider the Rift Valley as a linguistic area, the inclusion of further languages contributes to a better understanding of the presence (or absence) of regional (i.e. areal) features.

In Rangi, the verb-auxiliary construction is used to encode the immediate future tense and the general future tense, formed using the auxiliaries =**íise** and =**rɪ** respectively. In Mbugwe, the construction is found in the present imperfective, habitual, future perfective, and the past progressive. While Rangi also exhibits auxiliary-verb order in certain tense-aspect combinations, in Mbugwe all auxiliary constructions exhibit verb-auxiliary order in declarative main clauses. A summary of the relevant structures is shown below in Tables 15 and 16, along with examples from each language (see (50)–(54)). We follow the Bantu tradition and mark the auxiliaries with a hyphen (rather than explicitly as a clitic using =). However, for the purposes of the current study, we consider these constructions as analogous to those found in the other languages under examination in that they attract subject and tense-aspect information and form a sort of 'complex', albeit one that behaves somewhat differently from those from the other language families reported here.

(51)   *Mama óta árɪ maaji mpolɪ*
     mama   jót-a   **á-rɪ**   maaji   mpolɪ
     1.mother   get.water-FV SM1-AUX 6.water   later
     'Mother will get water later.' (Rangi; Gibson 2019: 763)

Table 15: Rangi verb-auxiliary constructions.

Form	Function
Verb SM-(TAM)-AUX	Immediate future tense **-íise**
	General future tense **-rɪ**

Table 16: Mbugwe verb-auxiliary constructions.

Form	Function
Verb SM=AUX	Present imperfective -*re*
	Habitual
	Past progressive
	Future perfective
	-re
	-je
	-áyse
	-áandá
	-jéénde
	-kɛɛndɛ

(52) *kilwire ɪkɪ kwanjula kirɪ*
       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.' (Rangi; Gibson 2019: 763)

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

(54) *osíra koje na vaána*
       o-sír-a   **ko-je**   na   va-ána
       INF-finish-FV   SM1PL-AUX.FUT1   CONN   2-child
       'We are going to die, and the children too.' (Mbugwe; Wilhelmsen 2018: 145)

In both Rangi and Mbugwe, the PCC is obligatory but only in certain tenses and also under certain syntactic constraints. In Rangi, the verb-auxiliary construction is obligatorily found in the immediate and general future tense (Dunham 2005; Gibson 2013, 2019; Stegen 2002). In Mbugwe, this is found in the present imperfective, habitual, future perfective, past progressive (53), future imperfective (54) (Mous 2000; Wilhelmsen 2014). For the current purposes we consider this to be obligatory since an attempt at a different word or omission of the cluster results in ungrammaticality. In terms of phonological characteristics, these comply with the broader phonotactics of the language. In other words, these units can be tone-bearing and in the case of the past-tense construction in Mbugwe for example, the auxiliary hosts the past tense prefix **á-** (cf. Wilhelmsen 2014: 3).

Table 17: Summary of KMN (2008). Types of historical explanation from Aikhenvald and Dixon (2001).

Language	PCC?	Historical explanation
Iraqw-Gorwaa	Yes	Retention (from Cushitic)
Alagwa	Yes	Retention (from Cushitic)
Burunge	Yes	Retention (from Cushitic)
Sandawe	Yes	Chance (possibly because of retention from Khoe-Kwadi)
Hadza	Yes	Unknown
Nyaturu	Yes	Borrowing or diffusion (from Proto West Rift)
Nyilamba-Ihanzu and Kimbu	Yes	Borrowing or diffusion (from Proto West Rift)
Rangi-Mbugwe	No	∅
Datooga	Yes	Borrowing or diffusion (from Proto West Rift)

4 Assessing the possibility of contact-based language change

In light of the data presented in the preceding sections, the current section seeks to assess the PCC found in each language in terms of its possible source. A summary of the historical explanation provided in KMN (2008) is presented in Table 17 below. To do so we take the analysis developed by KMN (2008) as a starting point and then using the further data now available to use, reconsider the historical explanation for the construction in each language in turn.

4.1 Nyaturu-Proto-West Rift

In KMN (2008), the preverbal clitic cluster in Nyaturu is described as arising from sustained contact with Proto-West Rift, the predecessor language to the current South Cushitic languages. Of the Nyaturu preverbal clitic cluster, it is written that "[i]t looks as if Bantu material has been used to build a system of preverbal clitics, encoding Bantu categories in a Southern Cushitic frame" (page 199). The possibility of the Nyaturu preverbal clitic cluster developing through entirely internal grammaticalisation processes (and therefore resembling the South Cushitic forms by chance) is ruled out because it cannot be traced to any auxiliary structures in nearby related languages, such as Sukuma and Nyamwezi.

What follows attempts to provide a bit more precision on KMN (2008)'s claim. Each morpheme of the Nyaturu preverbal clitic cluster will be considered, along with

a plausible source (or in some cases, two competing sources). The Nyaturu subordinate clause marker **nǐ**= (55) has two possible origins. The first is that it was borrowed into Nyaturu from Proto-West Rift or one of its predecessor languages – both Gorwaa and Iraqw use a preverbal clitic **ni**= to indicate a relative clause lacking an internal patient whose object is 1st person singular.

(55) *tí ni koom a paanga*

   tí   **ni**= ∅   **koóm** ~**LPA**~   ∅   **paanga -r´**
   PRO.DEM1.F   MP.S1=AUX   have.1SG ~REL~   AUX   machete-L.FR

   'this (thing) which I have is a machete' (Gorwaa; Harvey 2017: [20150808a.2])

This morpheme appears to be of good Cushitic origin, in that it appears with the same function in the geographically distant, but genetically-related Lowland East Cushitic language Harar Oromo (orm) as **-`n** (Owens 1985).

   Alternatively, and perhaps more plausibly, the Nyaturu subordinate clause marker may have its origins in a Bantu focus-marker, commonly of the form **ni** (cf. Schwarz 2003; Gibson 2019).

   The near past tense clitic **ájà** has many analogues in Bantu languages (c.f. **-zà** for perfective past disjoint in Sambaa (ksb) (Riedel 2009: 29), for example).

   The Nyaturu sequential morpheme **qàá** is quite possibly Bantu in origin as well. Masele indicates that the [ɣ] sound may have developed from a [g] sound (2001: 122) in Nyaturu. Though we do not have a morpheme **gàá** in any nearby language, Batibo (1985) identifies the Sukuma morpheme **-ka-** as "circumstantial" (page 259) and a "non-definitive or unaccomplished" (page 263) morpheme. Nurse (2000: 523) points out the **ka** of subsequent action, widespread in Bantu (but whose connection to the Nyaturu **qàá** would require some further phonological explanation (Nurse 2000: 533, ff11)).

   The Nyaturu persistive morpheme **kɨ̀ɨ** (56) also seems uncontroversially Bantu in origin, compare the Ihanzu auxiliary verb **kɨlɨ**, also used to convey a persistive meaning. Nurse (2000: 523) also draws attention to the widespread Bantu persistive aspect **ki**, but again notes that to connect this **ki** to the Nyaturu **kɨ̀ɨ** would require some further phonological explanation (Nurse 2000: 533, ff11).

(56) *ɨantʊ kɨlɨ ɨpegéha du ne*

   ɨ- a- ntʊ   **kɨlɨ**   ɨ- pegéh -a   du   ne
   AUG2-2-person   **still**   SM2-drill-FV   only   Q

   'Do people still drill?' (i.e. start fires by using fire drills) (Ihanzu; Harvey 2019b: [20201209_SKa.171])

The Nyaturu far past tense morpheme **náa** appears to be a good candidate for borrowing from Proto West Rift, or one of its predecessor languages (c.f. Iraqw/

Gorwaa "perfective" =**(g)a,** Alagwa "general past" =**áa**, and Burunge "preterite" =**áa**). For an alternative source, many Bantu languages employ a verb-internal morpheme *a*- to express past notions. For example, Sukuma employs a morpheme **a-** for "accomplished and immediative" (Batibo 1985: 263).

The Nyaturu near future tense morpheme **nàa** has also been proposed as a borrowing from Proto West Rift (c.f. Burunge "future1" =**aa**). With that said, because this form is only employed in Burunge, the direction of the transfer is still unclear (Roland Kießling, p.c. 08.09.2021).

The Nyaturu far future tense morpheme **ìkwí** was identified in KMN (2008: 199) as having no obvious origin either in Bantu nor in West Rift. We would submit that a possible Bantu source for this morpheme is the lexical verb "to stand" (Nyaturu **ɣw-ìmíka** c.f. Jinakìɨ̀ya Sukuma **gwìɨ̀ma** (Masele 2001: 594) and Rangi **kwìɨ̀mà** (Masele 2001: 709)). With that said, an anonymous reviewer points out that phonological form, as well as the semantics of "stand" or "stop" would not relate to future in any straightforward way, and overall, a closer analysis would need to be done to convincingly establish this link.

4.2 Datooga-PWR

KMN (2008) propose that speakers shifting from West-Rift to Pre-Datooga influenced the grammaticalization of Pre-Datooga constructions by incorporating Datooga morphology into West-Rift-like structures. KMN (2008) specifically mention three Datooga constructions, the future, persistive, and sequential, which resemble the South Cushitic clitic clusters in that they appear to be separable from lexical verbs to which they attach.

According to Datooga oral history (Wilson 1952), contact between speakers of Datooga and Iraqw is believed to have occurred in the TRV for at least the past 300 years, but little is known about contact between South Cushitic and Datooga groups further back in time. There is linguistic evidence of earlier periods of contact between West-Rift and Southern Nilotic speakers (Kießling 2002), which most likely occurred in areas north of the TRV. If the patterns exhibited by the Datooga future, persistive, and sequential constructions are due to contact which occurred within the TRV and did not involve speakers of other Southern Nilotic languages, then we would expect to see patterns which are not found in other Southern Nilotic languages.

There is evidence from the Kalenjin languages of the Southern Nilotic language which indicate that the persistive and sequential constructions have either been inherited or have developed independently of contact with South Cushitic languages.

Table 18: Southern Nilotic cognate persistive and sequential constructions.

	Datooga	**Nandi**	**Cherang'any**	**Akie**
Persistive	**(g)udu** (Gisamjanga), **ad-** (Asimjeeg)	**tà-**	**tʌ-**	**taa**
Sequential	**ǎg** (Gisamjanga), **V́-** (Asimjeeg)	**ak** 'and, with'	**ak=** 'and, with'	**ai, ka, anan** 'and, with'

This evidence comes in the form of cognate constructions (see Table 18), which resemble the Datooga constructions in both form and function (Creider and Creider 2001; Griscom 2019; König et al. 2015; Mietzner 2016).

The Kalenjin constructions which are cognate with the Datooga persistive construction are also used to code persistive aspect (often translated as 'still'). The persistive construction in Nandi (niq) and Cherang'any (enb) are analyzed as verbal prefixes, whereas the Akie (oki) persistive construction is analyzed as an auxiliary. The Nandi and Akie constructions which are cognate with the Datooga sequential constructions are analyzed as coordinating conjunctions, and the cognate construction in Cherang'any is analyzed as a verbal enclitic. The similarities between these constructions indicates that the Datooga persistive and sequential constructions were inherited from Proto-Southern Nilotic.

The diachronic source of the Datooga future construction is unknown (Roland Kießling p.c.). There are multiple cognate future constructions in the Kalenjin languages which can be linked to the verb **mac(-ey)** 'want' (e.g. Nandi **mâ-** and Akie **mach-**), and, although the verb **gas** 'want' in Datooga shares similar semantics with these Kalenjin verbs which serve as the source for the Kalenjin future constructions, it is not the source of the Datooga future construction. The development of a future tense construction from a verb meaning 'want' has been documented in many languages before (Heine and Kuteva 2005: 18), but there is no clear evidence linking the Datooga future construction specifically to a verb meaning 'want'.

There is, however, internal evidence indicating that the Datooga future and persistive constructions developed from multi-verbal constructions, in a parallel fashion to the future and persistive constructions of the Kalenjin languages. Kießling (2019) suggests that these Datooga constructions developed out of multi-verbal clauses and were simply inherited rather than borrowed or created as a result of contact.

One key feature of Datooga verbs which supports the putative verbal origins of the complex-stem constructions is the affirmative prefix **g-** ~ **q-**. Within the subject indexation paradigm of complex-stem constructions, third-person subject

indexation mandatorily co-occurs with a form that is identical to the affirmative prefix, as shown in (57). This form has been analyzed as either an instance of the affirmative prefix (Griscom 2019), as indicated in (57), or as part of the subject indexation prefix (Kießling 2000). Anderson (2011: 173) writes that the Datooga future construction may have developed from the fusion of a multi-verbal, doubly-inflected construction.

(57)   *i̧ndaw lapiyaedangw gigagonyi*
        ì̧ː-nd-áw   **lápíj-ɛɛ̀ː-d-àŋw**   **g-ì̧ː-g-à-gòɲí**
        COND-COP-POSS   money-PS.SG.SS.SG-2.SG.POSS   AFF-FUT-AFF-3-give:FS
        'If you have your money, he will give it (to you)' (Asimjeeg Datooga; Griscom 2018: [IGS0229_2016-12-12_#5_143])

The development of the doubling pattern could be described as having occurred in three stages (see Table 19). In the first stage, the two fully inflected verbs co-occur in a multi-verbal bi-clausal construction. In the second stage, the first verb grammaticalizes and in the process loses the capacity to co-occur with standard inflectional suffixes and a full subject indexation paradigm. The diachronic development of the second verb is less transparent, as it is tied to the wider distribution of the dependent stem structure in imperative and subordinate clause constructions, but it is clear that the second verb loses its capacity to co-occur with the affirmative prefix for non-3rd person subjects. In the third and final stage, the first verb becomes syntactically bound to the second verb and constitutes a cluster of preverbal morphology.

Kießling (2019) proposes a potential historical source for complex-stem constructions in Gisamjanga Datooga that consists of a verbal auxiliary followed by a lexical verb, corresponding to Stage II of the diachronic pathway. The co-occurrence of future + bound auxiliaries could be due to the merger of multiple verbs through the same pathway (i.e. FUT + AUX + VERB -> FUT=AUX=VERB), and some examples from the 1930s Berger corpus indicate that the future may have been more syntactically independent in the past (Kießling 2019). There is also evidence from Asimjeeg Datooga that the persistive may function independently as a lexical verb (58).

Table 19: Diachronic pathway for the development of complex-stem constructions in Datooga varieties.

Stage	Structure
I: Multi-verbal	AFF-SUBJ-V-suffixes AFF-SUBJ-V-suffixes
II: Auxiliary + lexical verb	AFF-(SUBJ)-AUX (AFF)-SUBJ-V-suffixes
III: Complex stem	AFF-AUX=(AFF)-SUBJ-V-suffixes

(58)   *mad akalaelae*
         m-àd   àkàlɛ́lɛ̀
         NEG-PERS   one
         'It's not one/It's not the same' (Asimjeeg Datooga; Griscom 2018: [2015-12-21_GG1_165])

4.3 Sandawe (coincidence)

KMN (2008) propose that any similarities between the Sandawe PGN clitics and the clitic constructions of the other TRV languages under discussion are most likely due to chance, and that the Sandawe PGN clitics are more than likely a retention, with similar structures found in languages of the Khoe-Kwadi family (with which Sandawe may have a distant genetic relationship). We find this claim to be supported by a number of pieces of evidence, some of which mentioned in KMN (2008), and some of which we will explore for the first time here.

KMN (2008) establishes three principal ways in which the Sandawe PGN clitics resemble preverbal clitic clusters of neighbouring languages: 1) Sandawe PGN clitics are non-verbal elements which mark for the subject, similar to preverbal clitic complexes in Southern Cushitic and Nyaturu; 2) "cliticization is to the left" (page 205); and 3) the Sandawe PGN clitic has a focus function, similar to preverbal clitic complexes in, for example, Burunge. Points (1) and (3) above are accurate, but the observation made in point (2) must surely be a mistake: Sandawe clitics are *enclitic* in nature, thus rendering these PCCs *different* to those found in South Cushitic and Ihanzu, Nyaturu, Nyilamba, but similar to those found in Hadza and (probably) Rangi-Mbugwe.

In terms of how the Sandawe PGN clitic differs from the preverbal clitic clusters of neighbouring languages, KMN (2008) notes that: 1) Sandawe PGN clitics do not mark tense-aspect, and 2) Sandawe PGN clitics do not have a fixed position. To this, we would add that 3) there exist no morphemes in the Sandawe PGN clitic paradigm which can be readily linked to any other language in the TRV, and that 4) Sandawe PGN clitics are monomorphemic, and are therefore not clusters of clitics at all.

Finally, it is worth revisiting a detail in KMN (2008: 205) noting that the Sandawe PGN clitics are similar in many respects to the PGN markers in the Khoe languages of southwest Africa. Syntactically, and as mentioned above, the PGN clitics of Sandawe are either enclitics to verbs (as in (42) and (44) above), or preverbal clitics encliticising to other constituents in the clause (as in (41) and (43) above). In the Khoe-Kwadi (and depending on the language), the PGN markers appear as nominal suffixes or as clitics on noun phrases, among other constituents (Witzlack-Makarevich and Nakagawa 2019: 397–398). In (59), an example from Namibian Khoekhoe (naq), the PGN marker =**b**, itself coding subject, encliticises to both the subject noun, *the boy* but also to the object noun *the girl*, which has its own PGN marker -*s*.

Table 20: Sandawe person-gender-number (PGN) clitics (column 2) compared with Sandawe personal pronouns (column 1) and PGN clitics in Proto-Khoe (column 3).

		Sandawe pers.pronouns		Sandawe PGN clitics (negative realis)		Proto-Khoe PGN system (Güldemann 2004: 297)		
		M	**F**	**M**	**F**	**Common**	**M**	**F**
Singular	1	*tʃí*		*sé*		**ti; *ta*		
	2	*hàpú*		*pó*			**tsa*	**sa*
	3	*hèwé*	*hèsú (hùsú)*	*é:*	*sú*		**bi*	**si*
Dual	1	—		—		**kho-mu*	**kho-mu*	**sa-mu*
	2					**kho-da-o*	**kho-da-o*	**sa-da-o*
	3					**kho-da*	**kho-da*	**sa-da*
Plural	1	*sű:*		*sű:*		**ta-e*	**!a-e*	**sa-e*
	2	*sĭ:*		*sĭ:*		**ta-o*	**!a-o*	**sa-o*
	3	*hèsó (hòsó)*		*só*		**nV*	**!a-u*	**di*

(59) /gôa**s**a**b** ge axa**b**a tsaurase go ‡gai
      /gôa-s-a=**b**   ge   axa-**b**-a   tsaurase go   ‡gai
      girl-3SG.F-OBJ=**3SG.M.SBJ** DECL boy-**3SG.M**-OBL gently REC.PST call
      'The boy called the girl gently.' (Namibian Khoekhoe; Haacke 2013: 329–330
      in Witzlack-Makarevich and Nakagawa 2019: 398)

Formally, KMN (2008: 205) note that the PGN markers "seem cognate" with those of the Khoe-Kwadi languages, and though there do not seem to be many resemblances between the PGN clitics of Sandawe and those reconstructed for Proto-Khoe (compare columns 2 and 3 in Table 20), Sandawe PGN clitics and personal pronouns certainly seem cognate (compare columns 1 and 2 in Table 20), a pattern which also obtains for the languages of Khoe-Kwadi.

As such, and with some further proof, it seems reasonable to assume that the PGN markers of Sandawe are not a result of areal contact, and that most (if not all) resemblance is due to chance. Whether the construction represents a retention specifically from Khoe-Kwadi, there is some evidence both syntactically and in terms of a similar grammaticalisation pathway to support this, though cognacy via any phonetically identifiable segments is not, at least on the surface, evident.

4.4 Hadza (unknown)

At the time during which KMN (2008) was written, the documentary and descriptive state of Hadza sufficed only for the article to confirm that Hadza featured a preverbal clitic cluster, but nothing more could be said about its historical development.

Advances in Hadza studies now allow us to make some observations. As mentioned above, the Hadza preverbal clitic complex is formed of a PGN-TAM marking element, as well as, in some cases, a set of auxiliaries expressing polarity, mood, and clause combining functions such as subordination and sequentiality. Our first observation deals with the auxiliaries, and the rest treat the PGN-TAM marking element.

Of the auxiliaries, the two marking sequentiality, **(h)a** and **ka**, are both candidates for cognacy with the "**-ka** type" sequentials mentioned elsewhere, such as the **qàá** sequential of Nyaturu (or even the **ka-** verbal morpheme of Sukuma) both described in Section 4.1 above.

Of the subject PGN-TAM marking element, there are several observations to make: the first will treat the TAM component, and the latter will treat the PGN component. Of the TAM component, the =**aa** enclitic for posterior tense (see Table 22) seems close enough to the West Rift =**aa** past enclitic to be worthy of note. With that said, it ought to be noted in Hadza that this form is phonetically [aʔa], whereas the West Rift form is phonetically [a:].

PGN-TAM marking in the veridical (expressing less certainty or counterfactual, see Table 21 below) is striking in that it is characterised by the morpheme **ikwi**.

Here there exist two possible explanations. The first suggests contact with Bantu (c.f. the form **ìkwí** identified in Nyaturu above, with an origin in the lexical verb "to stand"). The second suggests an internal origin: Hadza features a verb **ikha** defined in Miller et al. (2016) as "to stop (doing something)", or "to stand or to be standing". In fact, the meaning of the Hadza form is so close to the Bantu form, the Hadza verb itself may be a borrowing from Bantu. Whatever the case, the **ikwi**-future connection seems to be a common pattern across the languages in our sample.

Of the PGN component, the pronominal elements are almost always marked by /n/ in the 1st person, /t/ in the 2nd person, and /s/ in the 3rd person (see Table 22). A clear Hadza-internal origin for these person marking consonants is not entirely clear. The personal pronouns of Hadza feature /n/ in the 1st person (e.g. **ono** 'I') and /t/ in the 2nd person (e.g. **te** 'you'), but do not mark 3rd person with /s/. With this in mind, a Hadza-external origin for these forms could be entertained. In fact, this /n-t-s/ paradigm is strikingly similar to widespread patterns observed in the languages of the Afroasiatic phylum (c.f. Tucker 1967: 22). When looking for parallels within West Rift (Cushitic) – indeed the most likely group of languages to serve as the donor for

Table 21: Person-Gender-Number (PGN) markers in the veridical in Hadza.

	Singular	Plural
1st person	=**nikwi**	=**ukwi** (EXCL), =**bikwi** (INCL)
2nd person	=**tikwi**	=**ti:kwi**
3rd person	=**kwiso** (M), =**kwiko** (F)	=**kwisi** (M), =**kwise** (F)

Table 22: Person-gender-number (PGN) markers in the posterior, potential, veridical, and subjunctive in Hadza.

Posterior	Sg.	Pl.	Potential	Sg.	Pl.
1	=naa	Excl. =aa Incl. =baa	1	=nee	Excl. =ee Incl. =bee
2	=taa	M =(i)tia F =etea	2	=tee	M =itii F =etee
3	M =amo F =akwa	M =ami F =ame	3	M =heso F =heko	M =hisi F =hese
Veridical	**Sg.**	**Pl.**	**Subjunctive**	**Sg.**	**Pl.**
1	=nikwi	Excl. =ukwi Incl. =bikwi	1	=na	Excl. =ya Incl. =ba
2	=tikwi	=ti:kwi	2	=ta	M =si F =te
3	M =kwiso F =kwiko	M =kwisi F =kwise	3	M =so F =ko	M =si F =se

Afroasiatic pronominal material to Hadza – the evidence for this /n-t-s/ paradigm is not so convincing if we examine only the independent pronouns (Kießling and Mous (2003) reconstruct these as *ʔana 'I', *kii 'you (f.sg.)", *kuu 'you (m.sg.), and *ina 'he, she', with the 2nd person form lacking a /t/ morpheme, and the 3rd person form lacking an /s/ morpheme). Roland Kießling (p.c. 21/09/2022) notes that more promising evidence of the /n-t-s/ paradigm in West Rift can be found in the preverbal clitics marking subordinate clauses in Gorwaa and Iraqw, where the forms are **ni** 'that I', and **ta** 'that you' (see esp. Mous 1993: 125). Further, Kießling (2002: 358–360) reconstructs Proto-West Rift pronominal roots ***ʔani** (1.Sg) and ***ʔata** (2Sg) as the sources of these preverbal subordinate markers, which may themselves link to the Proto-East Cushitic pronominal roots **ani** (1.Sg) and **ata** (2Sg), reconstructed in Sasse (1981: 144). Kießling (p.c. 21/09/2022) further points out the Proto-West Rift 3rd person possessive pronominal suffix ***-s.**

Syntactically, it is worth noting that Hadza PGN-TAM marking clitics can not only occur before the verb (as enclitics to preverbal auxiliaries), but also occur as enclitics to the lexical verb. In this way, they are similar to the PNG-markers of Sandawe in that they are enclitic in nature.

Further, and as mentioned above, the subject marking of Hadza PGN enclitics can more-or-less be seen as deriving from pronominal material (in this case, material which seems Afroasiatic in origin). In this way too they are then similar to Sandawe PGN-TAM clitics, in that these are also historically linked to pronouns.

4.5 Rangi-Mubgwe

The compound auxiliary constructions found in Rangi and Mbugwe which we consider here to be part of the larger PCC constructions under examination in the present paper were not addressed by KMN (2008). The analysis developed here adopts however the same categories used in this section so far. There are no formal similarities between the Rangi-Mbugwe auxiliary forms and no reason to think that this represents a borrowing of form. However, there are functional similarities in that these constructions are associated with the encoding of specific tense-aspect-mood distinctions and also exhibit a sensitivity to clause type. The relevant socio-historical context involves a sustained history of contact between Rangi-Mbugwe speakers and Cushitic-speaking communities over an extended period of time. Well-established pathways of grammaticalisation can be identified which sees the development of auxiliary forms from (lexical) verbs, a process which is widespread across Bantu and cross-linguistically. The combination of auxiliary forms and main verbs to encode a wide range of tense-aspect distinctions has also been described across the language family.

5 Summary and conclusions

5.1 Assessing the pre-verbal clitic cluster as an areal feature

In KMN (2008), positing the pre-verbal clitic cluster as one of the 19 areal features makes two major underlying assumptions. The first is that the preverbal clitic clusters in the languages of the sample could be compared (which we will comment on in Section 5.1.1 below). The second was that so doing would help develop insights into i) the nature and mechanics of language contact (which we will comment on in Section 5.1.2 below), ii) the character of linguistic areas as a linguistic phenomenon, and iii) the history of the encounters between the peoples of the area (both of which we will address in Section 5.2 below).

5.1.1 Comparing the pre-verbal clitic clusters of the sample

Though KMN (2008) do not provide any criteria by which to identify or define a preverbal clitic cluster, the structures they describe can be characterised by several features: syntactic independence (i.e. the ability of the clitic cluster to occur separately from the verb, or to occur in more than one syntactic position), clustering of multiple morphemes (i.e. the clitic cluster is composed (either by agglutination or

fusion) of multiple identifiable morphemes), characteristic semantic domains (especially subject argument, object argument, case, tense, clause type, sequentiality and/or focus), and distribution (i.e. when the preverbal clitic cluster can occur,

Table 23: Overview of constructions examined in this paper.

Language	Syntactic independence	Cluster of multiple morphemes	Semantic domains	Distribution
Iraqw-Gorwaa	Yes	Yes	Mood, voice, core arguments, aspect, other	Obligatorily present in all finite clauses
Alagwa	Yes	Yes	Mood, clause type, core arguments, tense, other	Obligatorily present in all finite clauses, except if the subject is phonologically overt, in which case, clitic clusters marking only arguments may be omitted
Burunge	Yes	Yes	Mood, clause type, core arguments, tense/aktionsart; direction of action; other	Obligatorily present in all finite clauses
Nyaturu	Yes	Yes	Clause type, tense, subject, aspect/ clause-combining	Present only in certain tense/aspect constructions
Datooga	Bound to main verb, no intervening constituents	Yes	Polarity, tense, aspect	Present in all complex stem constructions
Ihanzu-Nyilamba	Yes	Yes	Clause type, negation, tense	Present only in certain tense/aspect, polarity, and clause-type combinationns
Sandawe	Bound to verb or other non-subject arguments, no intervening constituents	No	Person, gender, number, information structure	Present in all affirmative realis clauses which do not feature a subject-focus marker
Hadza	Can occur either independently or as an enclitic	Yes	Person, gender, number, tense, aspect	Present in all non-imperative verbal clauses
Rangi-Mbugwe	Yes	Yes	Tense-aspect, clause type	Present only in certain tense/aspect constructions

ranging logically from obligatory to restricted to only a few constructions). It should be noted that no one of these features (or absence thereof) is either necessary or sufficient to call a construction a preverbal clitic complex, but prototypical complexes tend to possess many of these features more or less robustly. Table 23 provides an overview of the constructions examined in the current paper.

5.1.2 The nature and mechanics of language contact (i.e. what borrowing a pre-verbal clitic complex means)

KMN (2008) makes clear that the most prototypical preverbal clitic clusters are those from the South Cushitic languages (Alagwa, Burunge, Gorwaa, and Iraqw), and indeed it is these languages (or their common predecessor language, Proto West-Rift) upon which most other languages of the Tanzanian Rift Valley Area (Nyaturu, Nyilamba-Ihaznu, and Datooga) are said to model their preverbal clitic complexes.

Indeed, the concept of 'model' is used advisedly here, in that it is essential to what it means to borrow a preverbal clitic cluster as a grammatical feature: linguistic material in language A forms the basis upon which linguistic material in language B is developed (c.f. Heine and Kuteva 2005). Crucially, this new linguistic material in language B need not be borrowed from language A. In the context of preverbal clitic complexes in the Tanzanian Rift Valley Area then, the occurrence of morphemes with cognates in South Cushitic, though a *sufficient* criterion for proposing a contact scenario, is not a *necessary* criterion for proposing a contact scenario. Indeed, the primary argument in KMN (2008) is that Rift Valley Area languages have adopted the South Cushitic preverbal clitic cluster as a frame, whose slots could then be filled with either inherited or borrowed morphemes.

5.2 Whither the Tanzanian Rift Valley (area)?

KMN (2008)'s presentation of preverbal clitic clusters in the languages of the Tanzanian Rift Valley was a subsection of a chapter with a larger goal: that is, the establishment of a 'Tanzanian Rift Valley Area' – the languages within which having come to resemble each other in at least 19 ways, due primarily to a situation of long-term, dynamic language contact. Subsection 5.2.1 provides an alternate approach, focusing on the history of the individual encounters between the peoples of the area. Subsection 5.2.2 reflects on what the analysis of preverbal clitic clusters put forth in this paper means in terms of the purported Tanzanian Rift Valley linguistic area and represents the conclusion of the paper.

5.2.1 Preverbal clitic clusters and the histories of the Tanzanian Rift Valley

In response to the problems entailed in establishing linguistic areas in any sort of coherent way, Campbell (2017: 34) describes such work as "not important nor practical", and concludes that the more fruitful inquiry lies in understanding the changes themselves (Campbell 2017: 34). We interpret this both in a (historical, socio-) linguistic sense, as well as a broader interdisciplinary sense. Below, we recapitulate the contact events suggested by our data on preverbal clitic clusters, provide some of the existing corroborating evidence (both linguistic and otherwise), and highlight what remains to be asked and understood about these cases of contact. In this way, we hope to indicate future directions of inquiry.

Nyaturu is probably the strongest case displaying development of a preverbal clitic cluster as a result of contact from South Cushitic. Two possible historical explanations are given for this in KMN (2008: 203) either i) "Nyaturu, or one form of Nyaturu, was once used by a group of bilingual West Rift [Cushitic] speakers", or "the Nyaturu, or a section of the Nyaturu, were once bilingual in West Rift [Cushitic] or one of its predecessors". It is generally accepted that the Nyaturu people were not linguistically homogeneous (see Section 2.3.2 above) nor politically centralised (e.g. Jellicoe 1969), and any approach to further understanding the linguistic history of Nyaturu-West Rift Cushitic contact would have to take this into account. Linguistic evidence of other contact-induced features posited as having developed in West-Rift Cushitic due to Bantu contact (two pasts and a subjunctive in **-ee**), as well as features posited to have developed in Pre-Burunge due to Bantu contact (a future tense and SVO argument order) are presented in KMN (2008). Additionally, Masele (2001: 395–397) lists a series of words in F-group Bantu languages thought to have come from Iraqw, though it is probably more appropriate to think of these as coming from (Proto-) West Rift. All of this is to say that contact between speakers of Nyaturu (or a predecessor language) and speakers of West Rift Cushitic (and descendent languages) is a well-motivated historical event, and further exploration of this historical contact (or contacts) between these groups could yield important insight.

Ihanzu-Nyilamba have less prototypical preverbal clitic complexes than Nyaturu, in that evidence for their syntactic independence is somewhat weaker, the semantic domains for which they mark less extensive, and the distribution rather more restricted. It is informative to compare the Ihanzu-Nyilamba preverbal clitic clusters with the Nyaturu preverbal clitic clusters, as there are important areas of overlap: of the four total clitic morphemes of the Ihanzu-Nyilamba preverbal clitic clusters (Nyaturu has 8), two (the subordinate **ni** and the "past 2" **áza**) also occur in Nyaturu. With that said, the Ihanzu-Nyilamba PCC system is not just an impoverished Nyaturu system, as one morpheme is entirely different (the Past3 **áli**), and one semantic category (negation) is not present in the Nyaturu system. This seems to

indicate that, while the formation of a preverbal clitic complex could have been two largely separate developments (arising once in Nyaturu and again in Ihanzu-Nyilamba), it is perhaps more likely to have been a shared process among these three languages (possibly initiated in a shared predecessor language). The morphemes shared between the Ihanzu-Nyilamba and Nyaturu preverbal clitic clusters could point to how the development of a preverbal clitic cluster began in a proposed Pre-Nyaturu-Ihanzu-Nyilamba (i.e. as a marker of clause type and tense), with the systems further developing after this predecessor language differentiated into Nyaturu and Ihanzu-Nyilamba. It also seems plausible that, after this separation, contact between Nyaturu and West Rift Cushitic remained strong, whereas contact between Ihanzu-Nyilamba and West Rift Cushitic was either less intense, or non-existent.

Datooga developing a preverbal clitic cluster as a result of contact with West Rift Cushitic seems, as a result of our analysis, to perhaps be a bit too late chronologically, as languages related to Datooga (but never demonstrably in contact with West Rift or any of its descendant languages) display similar constructions. As such, these may be inheritances from a predecessor language of Datooga (pre-Datooga, pre-Omotic-Datooga, or even pre-Southern Nilotic). As for the origin of these preverbal clitic clusters in the predecessor language, contact with a Cushitic language at an earlier point can't be ruled out (see Kießling and Mous 2003: 27), but this would require considerably more research (in a greater time-depth) to work out.

Analysis of Sandawe strengthens the argument that any similarity between South Cushitic and Sandawe is chance. Important differences from South Cushitic preverbal clitic clusters include that Sandawe preverbal clitic clusters are monomorphemic, enclitic elements, as well as that subject-marking is cognate with pronominal elements. Examination of specific similarities in Khoe-Kwadi languages also strengthens the argument that the Sandawe construction is a retention. In concluding her 1998 work evaluating links between Eastern and Southern "Khoisan" languages, Sands (1998: 166) concludes that while it is "a little more likely than not" that Sandawe and Southern Khoisan are related, "further research is needed to elucidate the relationship". The time-depth (and geographical distance) separating contemporary Sandawe and Khoe-Kwadi languages is massive, and distinguishing genuine cognate structures from chance resemblances would be a major task. Genetic and archaeological evidence would also play an important role.

Historically speaking, Hadza presents a perennial challenge in that it is most likely unrelated to any other language, and therefore it is especially difficult to determine whether any material has been borrowed into the language (Aikhenvald and Dixon's (2001) resemblance due to borrowing or diffusion), or whether it developed independently (resemblance due to chance). With that said, the preverbal clitic complex of Hadza is interesting in the sheer variety of Tanzanian Rift Valley languages which it resembles. Morphologically, there are forms in the Hadza

preverbal clitic cluster which resemble morphemes from both South Cushitic and Bantu languages. The forms are (like Sandawe), sometimes enclitic. Also like Sandawe, the subject-marking is cognate with pronominal elements *but*, this pronominal material has similar forms in Afroasiatic (either West Rift Cushitic or something older or different altogether). Again, we hesitate to label any of these "similar forms" cognates. Contact between Hadza and virtually all languages nearby (Datooga, Bantu varieties, as well as Iraqw and potentially earlier forms of South Cushitic) is evident in its lexicon (Miller et al. 2016). Senior (1938) records the visitation of large numbers of Sukuma-speaking people (estimated at some 30,000 people yearly) on the western shores of Lake Eyasi as part of an annual salt trade, and oral histories of the Ihanzu (Sanders 2008) and the Jinyakɨ‌ya Sukuma (Lusekelo 2021) saw communities of Hadzabe people as a refuge when crops failed during periodic drought. One genetic study (Tishkoff et al. 2007: 2191)[12] identifies a Hadza-Sandawe contact event some 15–20 thousand years ago, with little subsequent contact thereafter. Blurton Jones (2016: 18) correlates this to a "dry period [which] may have brought the Hadza and Sandawe together by degrading their preferred habitat and removing the montane forest that held them apart". Suffice it to say that the study of the history of the Hadza language is in its infancy, and that the Hadza preverbal clitic complex suggests rich layers of contact, some of which are possibly quite old.

5.2.2 Preverbal clitic clusters and the areal linguistics of the Tanzanian Rift Valley

In one sense, our analysis has confirmed what had been first asserted in KMN (2008): across the 12 Tanzanian Rift Valley languages of our sample, the preverbal clitic clusters are significantly similar in their syntactic, morphological, and semantic characteristics (see Section 5.1.1 above). They also fit with views of language contact and grammatical change described in Heine and Kuteva (2005) (see Section 5.1.2 above), in which a preverbal clitic cluster in one language serves as a model for the development of a preverbal clitic cluster in another, but which itself does not need to be borrowed as-is into that language to count as contact-induced change. Instead, and as described in KMN (2008), the preverbal clitic cluster spread as a more abstract concept: a syntactic/semantic frame whose slots were filled according to the linguistic resources a given language had at hand: a kind of grammatical *bricolage*. In this way, the preverbal clitic cluster is a strong candidate for an areal feature, and exemplifies an areal dynamic of language contact and change in a very convincing way.

---

12 Though we recognise the validity and utility of the genetic arguments made in Tishkoff et al. (2007), the interpretation of anthropological and linguistic data made therein is more controversial (c.f. Knisley 2021: 206–251 and "The Problem Space of Eastern African Khoisan Foragers").

In another sense, our analysis has complexified the story of the spread of pre-verbal clitic cluster. Firstly, KMN (2008) proposes West-Rift (Cushitic) as the model upon which other languages (Bantu, Datooga) of the area developed their preverbal clitic cluster. Though our analysis confirms that this appears to be the case with most of the languages of the sample (and, save Datooga, all of the languages for which there was sufficient data at the time of KMN (2008)), there exists another model in Sandawe which, with its freer syntactic distribution and subject-marking derived from pronouns, may have conceivably influenced the Hadza preverbal clitic cluster. In this way, even when considering one areal feature, the Tanzanian Rift Valley shows multilateral dynamics, rather than unilateral. Second, because the preverbal clitic cluster is not a unitary feature (i.e. its syntactic, morphological, semantic, and formal properties are not, even in the abstract, fixed), the contact-induced preverbal clitic clusters of the area have developed in various ways. For example, the preverbal clitic complex of Nyaturu encodes a rich array of grammatical information (clause type, tense, subject, aspect/clause-combining), whereas the (closely-related) Ihanzu-Nyilamba preverbal clitic complexes encode a more limited set of grammatical information (clause type, negation, tense). These patterns exhibit the tendencies of some Tanzanian Rift Valley languages to be more centrally involved in contact-induced changes (nuclear members), and others to be less so (peripheral members). Third, there is no indication that the development of the contact-induced preverbal clitic clusters took place at the same time: if Datooga developed its preverbal clitic cluster as a result of contact, this contact would probably have taken place at a much earlier (pre-Datooga, pre-Omotik-Datooga, or even pre-Southern Nilotic) stage, whereas the Nyaturu preverbal clitic complex probably began development at a much later date. In this way, Tanzanian Rift Valley languages display characteristics of contact which is chronologically layered.

In addition to this (and crucially for geographically-focused views of linguistic areas), and aside from the brief mention of Proto-Khoe in Section 4.3 above, no attempt has been made in this paper to determine whether preverbal clitic complexes, as defined here, exist in any of the languages outside of the area posited in KMN (2008). KMN (2008) itself notes that, of its control languages Swahili (swa), Oromo (orm), and (Kenyan) Maasai, none possess a preverbal clitic cluster. With that said, there exist a vast number of languages in the immediate vicinity of the Tanzanian Rift Valley for which very little, if any, documentation exists (including local Tanzanian varieties of Maasai), and one must wonder how the proposed linguistic area would look if this data were available.

This cursory evaluation of the Tanzanian Rift Valley Area is perhaps informative for individuals interested in linguistic areas: their conceptualisation, their delimitation, as well as their typologies. With that said, it is harder to establish how the Tanzanian Rift Valley Area – a geographically-centred entity – helps us understand the underlying histories, cultures, and, indeed, languages of the Tanzanian Rift

Valley. Indeed, rather than engaging in a discussion about whether our analysis militates for or against a Tanzanian Rift Valley Area, we see considerably more utility in employing our data to highlight the series of individual events that may have given rise to preverbal clitic clusters in the languages of our sample.

Abbreviations

Abbreviations follow the Leipzig Glossing Rules with the following additions:

1, 2, 3	Bantu noun classes 1, 2, 3 etc.
&	connective
[]GEN	tonal genitive
AFF	affirmative
AM	associated motion
ANA	anaphoric particle
ANT	anterior
AUG	augment
CONN	connective
CSEC	consecutive
DEM1, 2, 3	demonstrative of 1st, 2nd, 3rd-degree deixis
DO	direct object
EMPH	emphatic
FFUT	far future
FP, FPST	far past
FT	feminine t-type gender
FR	feminine r-type gender
FV	final vowel
HAB	habitual
IMPRF	imperfective
ITV	itive
L	linker
MO	masculine o-type gender
NC	narrative conjunction
OM	object marker
PC	(realis) pronominal clitic
PCC	preverbal clitic cluster
PGN	person-gender-number marker
PERS	persistive
POST	posterior
PRES	present
PRO	pronoun
PRST	persistive
PST1	"past 1"
PST2	"past 2"
PVC	preverbal clitic
SEQ	sequential
SF	subject focus
SM	subject marker
SP	specific
SS	secondary suffix

TERM	terminal applicative
$B	morphophonological operation in Gorwaa characterised largely by delabialisation or lowering of tone[13]

Appendix A: Orthographies

With the exception of Datooga, Sandawe, Rangi, and Mbugwe, which employ orthographies more or less consistent with the IPA, the remainder of the languages of the sample have their own orthographic conventions. The table below gives the correspondences between IPA symbols and how they are represented in the orthography of that language. It is important to note here that no language of our sample is regularly written, and most of the orthographies used in this paper have been developed by missionaries or linguists within the past 100 years. In this way, none of the orthographies are in any sense "official orthographies", "community orthographies", etc., but are best seen as working orthographies. Blocks in grey indicate that the IPA sound is not, for the language in question, a phoneme.

IPA	Alagwa, Burunge, Gorwaa, Iraqw	Ihanzu	Nyaturu	Hadza
[ɲ]	ny	ny	ny	ny
[ŋ]	ng	ng'	ng'	ng'
[ʔ]	'			'
[q']	q			
[ʃ]	sh	sh	sh	sh
[χ]	x			
[x]			gh	
[ħ]	hh			
[ʕ]	/			
[j]	y	y	y	y
[ɬ]	sl			sl
[tʃ]	ch	ch	ch	tc
[tʃʰ]				tch
[tʃ']				jj
[dʒ]	j	j	j	j
[dz]		z		z
[tsʰ]				tsh
[ts']	ts			zz
[tɬ]				tl
[tɬ']	tl			dl
[tɬʰ]				tlh
[kʷ]	kw	kw	kw	kw

---

13 This morphophonological operation $B, and its counterpart $A are explained in further detail in Harvey (2020).

(continued)

IPA	Alagwa, Burunge, Gorwaa, Iraqw	Ihanzu	Nyaturu	Hadza
[kʷʰ]				kwh
[gʷ]	gw	gw	gw	gw
[kʼ]				gg
[kʷʼ]				ggw
[ŋʷ]	ngw	ngw	ng'w	ng'w
[qʼʷ]	qw			
[χʷ]	xw			
[ɣ]			q	
[ʁ]			R	
[l]	l	l	l	r
[ǀ]				c
[ǀʰ]				ch
[ⁿǀ]				nc
[ⁿǀˀ]				cc
[ǃ]				q
[ǃʰ]				qh
[ⁿǃ]				nq
[ⁿǃˀ]				qq
[ǁ]				x
[ǁʰ]				xh
[ⁿǁ]				nx
[ⁿǁˀ]				xx
[á]	á	á	á	á
[aː]	aa	aa	aa	a:
[ɪ]		ɬ	ɬ	
[ʊ]		ʊ̵	ʊ̵	
[ɛ]		e	e	e
[ɔ]		o	o	o

Appendix B: Typological summaries of the preverbal clitic clusters of the Tanzanian Rift Valley

Alagwa

Criterion	Value	Notes
Syntactic independence	Yes	Intervening adverbs, and object Ns and NPs
Cluster or fusion of multiple morphemes	Yes	Mood, clause type, core arguments, tense, etc.

(continued)

Criterion	Value	Notes
Semantic domains		Mood (indicative, optative, consecutive, ventive, immediate), clause type (subordinate, impersonal subject), core arguments (subject, object), tense (general past, perfect, predicative focus), other (beneficient, ablative, applicative, instrumental)
Distribution		Obligatorily present in all finite clauses, except if the subject is phonologically overt, in which case clitic clusters which mark only the arguments may be omitted
Host of Cliticisation		– as per Harvey (2018, specifically pp. 137–162) analysis: a semantically null and phonetically null auxiliary ∅ – as per Mous (2016, see especially 1993: 173–192) analysis: forms meaning roughly "to be"
Direction of Cliticisation		Both proclitics (e.g. mood, clause type, core arguments) and enclitics (e.g. tense, other)

Burunge

Criterion	Value	Notes
Syntactic independence	Yes	Intervening adverbs, and object Ns and NPs
Cluster or fusion of multiple morphemes	Yes	Mood, clause type, core arguments, tense/aktionsart, direction of action, etc.
Semantic domains		Mood (indicative, conditional, optative), clause type (subject focus, object relative, indefinite subject, benefactive), core arguments (subject, object), tense/aktionsart (present, preterite, completive, future 1, future 2, habitual, prospective), direction of action (ventive, reflexive, separative), other (object focus, comitative/instrumental, sequential)
Distribution		Obligatorily present in all finite clauses
Host of Cliticisation		– as per Harvey (2018, specifically pp. 137–162) analysis: a semantically null (and sometimes also phonetically null) auxiliary ∅ or **i** – as per Mous (2016, see especially 1993: 173–192) analysis: forms meaning roughly "to be"
Direction of Cliticisation		Both proclitics (e.g. mood, clause type, core arguments) and enclitics (e.g. tense/aktionsart, direction of action, other)

Datooga

Criterion	Value	Notes
Syntactic independence	Bound to main verb, no intervening constituents	
Cluster or fusion of multiple morphemes	Yes	
Semantic domains	Polarity, tense, aspect	
Distribution	All "complex-stem" constructions	
Host of Cliticisation	Future relative morpheme of the form *dʒá* (or similar)	
Direction of Cliticisation	Proclitics	

Hadza

Criterion	Value	Notes
Syntactic independence	Can occur either independently or as an enclitic	
Cluster or fusion of multiple morphemes	Yes	
Semantic domains	Person, gender, number, tense, aspect	
Distribution	In all non-imperative verbal clauses	
Host of Cliticisation	– the lexical verb; or – preverbally, the auxiliary or an adverb; or – preverbally, the PGN marker may be syntactically independent	
Direction of Cliticisation	Enclitics	

Iraqw-Gorwaa

Criterion	Value	Notes
Syntactic independence	Yes	Intervening adverbs, and object Ns and NPs
Cluster or fusion of multiple morphemes	Yes	Mood, voice, core arguments, aspect, etc.
Semantic domains	Mood (indicative, conditional, prohibitive, questioning), voice (active, mediopassive), core arguments (sole argument of intransitive, agent and patient of transitive), aspect (perfective, imperfective, expectational, consecutive, background), other (reason, lative, ablative, instrumental)	
Distribution	Obligatorily present in all finite clauses	
Host of Cliticisation	– Harvey (2018, specifically pp. 137–162) analysis: a semantically null (and sometimes also phonetically null) auxiliary ∅ or **a** – Mous (1993, 2005, see especially 1993: 123–154) analysis: forms meaning roughly "to be"	
Direction of Cliticisation	Both proclitics (e.g. mood, core arguments) and enclitics (e.g. aspect, other)	

Ihanzu-Nyilamba

Criterion	Value	Notes
Syntactic independence	Yes	Relative ordering of negation and tense clitics is not fixed
Cluster or fusion of multiple morphemes	Yes	Clause type, negation, tense
Semantic domains		Clause type (main, subordinate), negation, tense (past 2, past 3)
Distribution		Present only in certain tense/aspect, polarity, and clause-type combinations (see row above)
Host of Cliticisation		The lexical verb
Direction of Cliticisation		Proclitics

Nyaturu

Criterion	Value	Notes
Syntactic independence	Yes	Intervening subject of an intransitive verb
Cluster or fusion of multiple morphemes	Yes	Clause type, tense, subject, aspect/clause-combining
Semantic domains		Clause type (main, subordinate), tense (far past, near past, near future, far future), subject, aspect/clause-combining (sequential, persistive)
Distribution		Present only in certain tense/aspect constructions (see row above)
Host of Cliticisation		– the lexical verb; or – an auxiliary verb
Direction of Cliticisation		Proclitics

Rangi-Mbugwe

Criterion	Value	Notes
Syntactic independence	Yes	But note, no element can intervene between the verb and the auxiliary
Cluster or fusion of multiple morphemes	Yes	Distinct auxiliary forms can be identified but this combines with subject information, and TA information (in Mbugwe). No examples of auxiliary hosting object marking.
Semantic domains		Specific tense-aspect combinations (Rangi: immediate future tense, general future tense. Mbugwe: Present imperfective, habitual, past progressive, future perfective). Found also in negation, relative clauses, interrogatives and subordinate clauses although these exhibit auxiliary-verb order in both languages.
Distribution		Present only in certain tense/aspect constructions (see row above)

(continued)

Criterion	Value Notes
Host of Cliticisation	An auxiliary verb (which, in all cases except for Rangi immediate future and far future tenses) is post-verbal
Direction of Cliticisation	Proclitics

Sandawe

Criterion	Value	Notes
Syntactic independence	Bound to verb or other non-subject arguments, with no intervening constituents	
Cluster or fusion of multiple morphemes	No	
Semantic domains	Person, gender, number, information structure	
Distribution	Present in all affirmative realis clauses which do not feature a subject-focus marker.	
Host of Cliticisation	– the lexical verb, and/or – a non-Subject clause constituent (e.g. and object or adverb)	
Direction of Cliticisation	Enclitics	

Appendix C: Evaluations of preverbal clitic complexes in KMN, versus in the current paper

Datooga

KMN (2008)	
Analysis	Contact-induced change, or more specifically change through language shift
Formal similarities	Cluster of morphemes
Functional similarities	Tense, aspect, polarity
Sociohistorical context	Historical contact between Datooga and Iraqw communities in past 300 years, but before that unknown.
Plausible grammaticalization pathways	V + V -> AUX V

Current paper	
Analysis	Inherited construction

Hadza

KMN (2008)	
Analysis	Not enough evidence
Formal similarities	Forms non-Bantu
Functional similarities	
Sociohistorical context	
Plausible grammaticalization pathways	

Current paper	
Analysis	– Auxiliaries *(h)a* and *ka* (sequential) resemble Bantu forms – *ikwi* (veridical) TAM marker resembles (possibly) Bantu forms "stand", or a far future morpheme – *aa* (posterior) TAM marker resembles Proto West Rift form for past – Subject indexation morphemes /n/ for 1st person, /t/ for 2nd person, and /s/ for 3rd person resemble the personal pronouns in Hadza for 1st and 2nd person, but a more precise fit is forms from Afroasiatic, though not necessarily from Proto West Rift – Syntactically, forms are enclitics, similar to Sandawe PGN-TAM enclitics – Subject-marking derives from (Afroasiatic) pronominal morphemes, similar to subject-marking in Sandawe PGN-TAM enclitics (which derive from Sandawe pronominal morphemes)

Nyaturu

KMN (2008)	
Analysis	Diffusion from South Cushitic (West Rift)
Formal similarities	The forms **náa**, **nàa**, and **ìkwɨ́** have no identifiable Bantu source. Forms *náa*, and *nàa* are probably from West Rift Cushitic. The source for the form **ìkwɨ́** is unknown.
Functional similarities	Uses as markers of subordination, tense, and sequentiality are similar to South Cushitic
Sociohistorical context	Ancestors of Nyaturu-speakers, speaking a language more like contemporary Sukuma/Nyamwezi, moved into a West Rift-speaking area and interacted over hundreds of years. This innovation was produced either by speakers of West Rift bilingual in pre-Nyaturu, or by speakers of pre-Nyaturu bilingual in West Rift.
Plausible grammaticalization pathways	Sequential/Persistive: verb-internal inflectional morpheme > pre-verbal auxiliary Tense/Aspect: borrowing Relative: ?

Current paper	
Analysis	The form **ìkwɨ́** has a possible Bantu source. The forms *qàá* and *kɨ̀* also have plausible Bantu origins.

Rangi-Mbugwe

KMN (2008)	
Analysis	Not considered in KMN (2008)
Formal similarities	The forms are all clearly of Bantu (verbal) origin and there is no reason to suggest that the forms themselves represent any kind of borrowing or are the result of contact.
Functional similarities	Tense-aspect-mood domain and sensitivity to clause type.
Sociohistorical context	Rangi and Mbugwe-speaking communities have been in sustained contact with Southern Cushitic (and to a lesser extent Nilotic) communities over time. Difficult to say at what point these structures arose however contact may have played a role in the rise of the typologically and comparatively unusual verb-auxiliary construction found in these languages
Plausible grammaticalization pathways	Common pathway of change involving the grammaticalisation of verb forms into auxiliaries (and in some instances subsequently into TA markers). V + V -> AUX V -> cleft-V + Aux > V Aux

Current paper	
Analysis	Independent innovation although perhaps facilitated by the history of contact in the area (cf. Gibson and Marten, Gibson 2019).

Sandawe

KMN (2008)	
Analysis	Coincidence
Formal similarities	None
Functional similarities	Information structure
Sociohistorical context	Potential pre-historical contact between PWR and Pre-Sandawe.
Plausible grammaticalization pathways	Subject pronoun -> subject enclitic

Current paper	
Analysis	Inherited construction

References

Aikhenvald, Alexandra Y. & Robert M. W. Dixon (eds.). 2001. *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press.

Anderson, Gregory. 2011. Auxiliary verb constructions in the languages of Africa. *Studies in African Linguistics* 40(1–2). 1–409.

Bala, Gudo G. (Bonny Sands & Will Grundy eds.). 1998. *Hadza stories and songs*. Los Angeles: Friends of the Hadzabe.

Batibo, Herman. 1985. *Le Kesukuma: Langue Bantu de Tanzanie*. Paris: Editions Recherche sur les Civilisations.

Blurton Jones, Nicholas. 2016. Why do so few Hadza farm? In Brian F. Codding & Karen L. Kramer (eds.), *Why forage? Hunters and gatherers in the twenty-first century* (School for Advanced Research Advanced Seminar Series), 113–136. Santa Fe: School for Advanced Research Press.

Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press.

Creider, Jane T. & Chet A. Creider. 2001. *A dictionary of the Nandi language*. Köln: Köppe.

Dryer, Matthew. 1991. SVO languages and the OV: VO typology. *Journal of Linguistics* 27(2). 443–482.

Dunham, Margaret. 2005. *Éléments de description du langi, langue bantu F.33 de Tanzanie*. Leuven, Belgium: Éditions Peeters.

Eaton, Helen. 2008. *Sandawe grammar*. SIL International.

Edenmyr, Niklas. 2004. The semantics of Hadza gender assignment: A few notes from the field. *Africa & Asia* 4. 3–19.

Gibson, Hannah. 2012. *Auxiliary placement in Rangi: A dynamic syntax perspective*. London: University of London Dissertation, ms.

Gibson, Hannah. 2013. Auxiliary placement in Rangi: A case of contact-induced change? *SOAS Working Papers in Linguistics* 153–166.

Gibson, Hannah. 2019. The grammaticalisation of verb-auxiliary order in East African Bantu. *Studies in Language* 43(4). 757–799.

Griscom, Richard T. 2018. Documentation of Isimjeeg Datooga. Endangered languages archive. Available at: http://hdl.handle.net/2196/00-0000-0000-000E-D158-9.

Griscom, Richard T. 2019. *Topics in Asimjeeg Datooga verbal morphosyntax*. Eugene, Oregon: University of Oregon PhD Dissertation.

Griscom, Richard & Andrew Harvey. 2020. Hadza: An archive of language and cultural material from the Hadzabe people of Eyasi (Arusha, Manyara, Singida, and Simiyu regions, Tanzania). Endangered languages archive. Handle. Available at: http://hdl.handle.net/2196/82e2b99d-5c62-4210-8903-8dd976337c10.

Güldemann, Tom. 1999. The genesis of verbal negation in Bantu and its dependency on functional features of clause types. In Jean-Marie Hombert & Larry M. Hyman (eds.), *Bantu historical linguistics: Theoretical and empirical perspectives*, 547–585. Stanford: Center for the Study of Language and Information (CSLI).

Güldemann, Tom. 2004. Reconstruction through 'de-construction': The marking of person, gender, and number in the Khoe family and Kwadi. *Diachronica* 21. 251–306.

Guthrie, Malcolm. 1967–1971. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages*. Farnborough: Gregg Press.

Haacke, Wilfrid H. G. 2013. Namibian Khoekhoe (Nama/Damara). In Vossen (ed.), *The Khoesan Languages*, 141–151. Routledge.

Harvey, Andrew. 2017. Gorwaa: An archive of language and cultural material from the Gorwaa people of Babati (Manyara Region, Tanzania). Endangered languages archive. Available at: http://hdl.handle.net/2196/00-0000-0000-000F-79D0-1.

Harvey, Andrew. 2018. *The Gorwaa noun: Toward a description of the Gorwaa language*. London: SOAS PhD Dissertation.

Harvey, Andrew. 2019a. Gorwaa (Tanzania) – language contexts. *Language Documentation and Description* 16. 127–168.

Harvey, Andrew. 2019b. Ihanzu: An archive of language and cultural material from the Ihanzu people of Mkalama (Singida Region, Tanzania). Endangered languages archive. Available at: http://hdl.handle. net/2196/00-0000-0000-0014-1365-F.

Harvey, Andrew. 2020. Verbal paradigms in Gorwaa: Phonological analysis in service of a unified account. Talk given at the Rift Valley Network Webinar Series. 06/05/2020.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Jellicoe, Marguerite. 1969. The Turu resistance movement. *Tanganyika notes and records* 70. 1–12.

Johnson, Frederick. 1923. Notes on Kiniramba. *Bantu Studies* 2. 167–192/223–268.

Kagwema, Augustino. 2020. Kikimbu: Documenting nomadism in central Tanzania. Endangered languages archive. Available at: http://hdl.handle.net/2196/9a577657-441e-40c1-8087-16debd971b10.

Kießling, Roland. 1994. *Eine Grammatik des Burunge*. Hamburg: Research and Progress.

Kießling, Roland. 2000. Verb classes in Nilotic: Evidence from Datooga (southern Nilotic). In Ekkehard Wolff & Orin David Gensler (eds.), *Proceedings of the 2nd world congress of African linguistics, Leipzig 1997*. Cologne: Rüdiger Köppe Verlag.

Kießling, Roland. 2002. *Die rekonstruktion der südkuschitishen Sprachen (West-Rift): von den systemlinguistischen Manifestationen zum gesellscaftlichen Rahmen des Sprachwandels*. Köln: Rüdiger Köppe Verlag.

Kießling, Roland. 2007. Alagwa functional sentence perspective and "incorporation". In Amha Azeb, Maarten Mous & Graziano Savà (eds.), *Omotic and Cushitic studies: Papers from the 4th Cushitic omotic conference, Leiden, 10–12 April 2003*. Köln: Rüdiger Köppe Verlag.

Kießling, Roland. 2019. Tracking down (and making sense of) the associated motion component in the Datooga lexicon. Presented at the East Africa Day at Leiden University, Leiden University.

Kießling, Roland & Maarten Mous. 2003. *The lexical reconstruction of West-Rift Southern Cushitic* (Kuschitische Sprachstudien 21). Köln: Rüdiger Köppe Verlag.

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

Knisley, Matthew. 2021. *Historical landscapes of the Sandawe Homeland, North-Central Tanzania*. University of Chicago Doctoral dissertation.

König, Christa, Bernd Heine & Karsten Legère. 2015. Discourse markers in Akie, a Southern Nilotic language of Tanzania. In Osamu Hieda (ed.), *Information structure and Nilotic languages*. Tokyo: Tokyo University of Foreign Studies.

Languages of Tanzania Project [LOT]. 2009. *Atlasi ya Lugha za Tanzania [Language Atlas of Tanzania]*. Dar es Salaam: Mradi wa Lugha za Tanzania, Chuo Kikuu cha Dar es Salaam.

Lewis, Paul M., Gary F. Simons & Charles D. Fennig (eds.). 2013. *Ethnologue: Languages of the world*, 17th edn. Dallas, TX: SIL International.

Lusekelo, Amani. 2021. Plant nomenclature and ethnobotany of the Hadzabe Society of Tanzania. Talk Given at Rift Valley Webinar Series 24/02/2021.

Maho, Jouni Filip. 2009. *New updated Guthrie list*. Brill.

Masele, Balla F. Y. P. 2001. *The linguistic history of Sisuumbwa, Kisukuma, and Kinyamweezi in Bantu Zone*. Memorial University of Newfoundland, St. John's PhD Dissertation.

Mietzner, Angelika. 2016. *Cherang'any, a Kalenjin language of Kenya*. Köln: Rüdiger Köppe.

Miller, Kirk, Mariamu Anyawire, Gudo G. Bala & Bonny Sands. 2016. *A Hadza lexicon*. ms.

Mitchell, Alice. 2021. Phasal polarity in Barabaiga and Gisamjanga Datooga (Nilotic): Interactions with tense, aspect, and participant expectation. In Raija Kramer (ed.), *The expression of phasal polarity in African languages*, 419–442. Berlin: De Gruyter Mouton.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60. 847–893.

Mous, Maarten. 1993. *A grammar of Iraqw*. Hamburg: Helmut Buske Verlag.

Mous, Maarten. 1994. Ma'a or Mbugu. In Peter Bakker & Maarten Mous (eds.), *Language intertwining*, 175–200. Amsterdam: IFOTT.

Mous, Maarten. 2000. Counter-universal rise of infinitive-auxiliary order in Mbugwe (Tanzania, Bantu F34). In Rainer Vossen, Angelika Mietzner & Antje Meissner (eds.), *Mehr als nur Worte …: Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland Festschrift*, xxx–xxx. Köln: Rüdiger Köppe Verlag.

Mous, Maarten. 2001. Basic Alagwa syntax. In Andrzej Zaborski (ed.), *New data and new methods in Afroasiatic linguistics: Robert Hetzron in memoriam*, 125–135. Wiesbaden: Otto Harrassowitz Verlag.

Mous, Maarten. 2005. Selectors in Cushitic. In Erhard Voeltz (ed.), *Studies in African linguistic typology*, 303–325. Amsterdam: John Benjamins.

Mous, Maarten. 2007. *A sketch of Iraqw grammar*. (unpublished manuscript). Leiden University.

Mous, Maarten. 2016. *Alagwa: A South Cushitic language of Tanzania: Grammar, texts and lexicon*. Köln: Rüdiger Köppe Verlag.

Mous, Maarten. 2021. Towards the linguistic history of Rangi and Mbugwe. Talk Given at Rift Valley Webinar Series 27/01/2021.

Nurse, Derek. 2000. Diachronic morphosyntactic change in western Tanzania. In Rainer Voßen, Angelika Mietzner & Antje Meißner (eds.), *Mehr als nur Worte: Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland*, 517–534. Köln: Rüdiger Köppe Verlag.

Nurse, Derek & Gérard Philippson. 2003. Towards a historical classification of the Bantu languages. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 164–181. London: Routledge.

Olson, Howard S. 1964. *The phonology and morphology of Rimi*. [Hartford Studies in Linguistics 14]. Hartford, Connecticut: Hartford Seminary Foundation.

Owens, Jonathan. 1985. A grammar of Harar Oromo (Northeastern Ethiopia). In *Kuschitische Sprachstudien, Cushitic language studies, 4*. Hamburg: Helmut Buske Verlag.

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

Rottland, Franz. 1982. *Die südnilotischen Sprachen: Beschreibung, Vergleich und Rekonstruktion*. Berlin: Reimer.

Sanders, Todd. 2008. *Beyond bodies: Rainmaking and sense making in Tanzania*. Toronto: University of Toronto Press.

Sands, Bonny. 1995. *Evaluating claims of distant linguistic relationships: The case of Khoisan*. Los Angeles: UCLA PhD Dissertation.

Sands, Bonny. 1998. *Eastern and Southern African Khoisan: Evaluating claims of distant linguistic relationships*. Köln: Rüdiger Köppe Verlag.

Sands, Bonny. 2013a. Morphology: Hadza. In Rainer Vossen (ed.), *The Khoesan languages*, 107–123. New York: Routledge.

Sands, Bonny. 2013b. Phonetics and phonology: Hadza. In Rainer Vossen (ed.), *The Khoesan languages*, 38–42. New York: Routledge.

Sands, Bonny. 2013c. Syntax: Hadza. In Rainer Vossen (ed.), *The Khoesan languages*, 265–274. New York: Routledge.

Sasse, Hans-Jürgen. 1981. Afroasiatisch. In Thilo Schadeberg (ed.) *Die Sprachen Afrikas*, Band 2, 129–148. Hamburg: Buske.

Schubert, Ralph, Anette Schubert, Douglass Boone & Sheri Daggett. 1997. Datooga dialect Survey. Manuscript.

Schwarz, Florian. 2003. *Focus marking in Kikuyu*. Berlin: Humboldt University MA dissertation.

Senior, H. S. 1938. Sukuma salt caravans to Lake Eyasi. *Tanganyika Notes and Records* 6. 87–90.

Steeman, Sander. 2012. *A grammar of Sandawe: A Khoisan language of Tanzania*. Leiden: Leiden University PhD Dissertation. Utrecht: LOT.

Stegen, Oliver. 2002. Derivational processes in Rangi. *Studies in African Linguistics* 31. 129–153.

Stegen, Oliver. 2003. First steps in reconstructing Rangi language history. Presented at the 33rd Colloquium on African Languages and Linguistics, Leiden University.

Tishkoff, Sara A., Mary Katherine Gonder, Brenna M. Henn, Holly Mortensen, Alec Knight, Christopher Gignoux, Neil Fernandopulle, Godfrey Lema, Thomas B. Nyambo, Uma Ramakrishnan, Floyd A. Reed & Joanna L. Mountain. 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution* 24(10). 2180–2195.

Tucker, Archibald Norman. 1967. Erythraic elements and patterning: Some East African findings. *African Language Review* 6. 17–25.

Whiteley, Wilfred H. 1958. *A short description of item categories in Iraqw, with material on Gorowa, Alagwa, and Burunge*. Kampala: East African Institute of Social Research (EAISR).

Wilhelmsen, Vera. 2014. *Periphrastic verbs in Mbugwe (F34)*. Uppsala: Uppsala University.

Wilhelmsen, Vera. 2018. *A linguistic description of Mbugwe with focus on tone and verbal morphology*. Uppsala University PhD Dissertation, ms.

Wilson, G. McL. 1952. The Tatoga of Tanganyika Pt. 1. *Tanganyika Notes and Records* 33. 35–47.

Witzlack-Makarevich, Alena & Hirosi Nakagawa. 2019. Linguistic eatures and typologies in languages commonly referred to as 'Khoisan'. In Ekkehard H. Wolff (ed.), *The Cambridge handbook of African linguistics*, 382–416. Cambridge: Cambridge University Press.