Chapter 13

The pragmatics of Swahili relative clauses

Mohamed Mwamzandi

University of North Carolina at Chapel Hill

Several studies explain the variation of the Swahili relative clause (RC) from a syntactic perspective. These studies discuss the derivational and structural differences/similarities between the *amba* RC and the tensed RC. In this study, the choice between the *amba* and tensed RCs is explained from a pragmatic perspective. 440 RCs were extracted from the Helsinki Corpus of Swahili. The dataset was then coded for various variables including relative marker (*amba*/tensed), relative type (restrictive/non-restrictive), length (number of words used), and information status (topic/non-topic). The results show that the tensed RCs are mostly restrictive while the *amba* RCs are mostly non-restrictive. Further, the mean length of the *amba* RC is higher than of the tensed RC. It was observed that the *amba* RC is preferred in topic shift transition, that is, when a non-topic NP becomes the topic NP in the following utterance while the tensed RC is preferred in continue transition, that is, if the topic of the matrix clause is the same as that of the RC.

1 Introduction

Several studies discuss the syntax of the Swahili tensed relative marker (RM) (1) and the *amba* RM (2). The goal of these studies is to draw a parallel between the two by explaining how the position of the RM is derivable from the same underlying position (Vitale 1981, Keach 1985) or occupies the same syntactic position (Demuth & Harford 1999, Ngonyani 2001, 2006). The examples presented in this study are from the Helsinki Corpus of Swahili (written texts in the standard Swahili), otherwise, citation will be given. The following notational conventions are followed: the RM is glossed as SRM (subject RM) or ORM (object RM) and the RC is bracketed.

(1) Askari   [a-li-ye-ki-ongoza   kikosi].
    1.policeman 1.SM-PST-1.SRM-7.OM-lead 7.squad]
    'The policeman who led the squad.'

(2) Askari   [amba-ye a-li-ki-ongoza   kikosi].
    1.policeman amba-1.SRM 1.SM-PST-7.OM-lead 7.squad
    'The policeman who led the squad.'

In (1) the SRM *ye-*, which agrees with the head noun *askari* 'policeman' in number and noun class is prefixed to the verb after tense while in (2) it is suffixed to *amba*. In Swahili and other Bantu languages, a predicate has a subject marker prefix that agrees with the subject in number and noun class and may also carry an object marker prefix. In (1), the subject marker prefix *a-* occurring in the verb initial position corresponds to the noun class 1 NP *askari* 'police' while the object marker prefix *ki-* occurring after the SRM *ye-* in (1) and after the tense marker *li-* in (2) corresponds to the noun class 7 NP *kikosi* 'squad'. I should mention that there is yet another type of RM referred to as the general relative exemplified in (3).

(3) Askari   a-ki-ongoz-a-ye   kikosi.
    1.policeman 1.SM-7.OM-lead-FV-1.SRM squad
    'The policeman who is leading the squad.'

The SRM *-ye* in (3) is suffixed to the verb after the final vowel. Time is unspecified in the general relative and it is used in cases interpreted as present or habitual (Ashton 1944). In this study, I explain the variation of the tensed RC and the *amba* RC from a pragmatic perspective and leave the general relative for future research. "Pragmatics" is defined as "the systematic study of meaning by virtue of, or dependent on, the use of language" (Huang 2007: 2). Notice that the tensed and the *amba* relative clause in (1) and (2) are grammatical and are truth conditionally equivalent. The aim of this study is to give a pragmatic account of the variation. I should mention that most of the explanations for the variation of the Swahili *amba* and tensed RCs in this study are not absolute but statistical tendencies and should be considered in their entirety rather than individually.

Based on a few selected examples from texts, earlier studies have made claims that the *amba* RM is required in non-restrictive RCs, but this is yet to be confirmed using a large dataset (Ashton 1944, Schadeberg 1989). A restrictive RC as illustrated in (5) delimits the referent of an NP (Andrews 2007: 206). On the other hand, a non-restrictive RC as illustrated in (4) makes a comment about an NP or another constituent without delimiting its referent.

(4) The Japanese [who/that are industrious] now outcompete Europe.

(5) The Japanese, [who/(*that) are industrious], now outcompete Europe. (Keenan 1985: 168)

In English, the restrictive clause in (4) allows the occurrence of the complementizer *that* but the non-restrictive clause in (5) does not. It is also possible to suppress the relative pronoun in a restrictive RC as in in: *The man I saw yesterday left this morning*, but not in a non-restrictive RC (Comrie 1989: 139). Further, as indicated by the comma before and after the RC, the non-restrictive clause as seen in (5) is set off intonationally from the main clause. Although the dataset in this study shows that the *amba* relative is mostly used in non-restrictive RCs, both the tensed and *amba* RCs can be used restrictively as well as non-restrictively. Thus, the restrictive and non-restrictive variable alone may not explain the Swahili *amba* and tensed RC variation. In addition to being used in non-restrictive RCs, it has been claimed that the *amba* RC when contrasted with the tensed RC is morphologically more versatile; its word order more flexible; and its structure more complex (Schadeberg 1989, Russell 1992).

I explore via corpus analysis the following questions:

1. The frequency of use of the *amba* RC and tensed RC as restrictive/non-restrictive and its implications in the variation.

2. Whether the speaker's choice of the *amba*/tensed RC is influenced by the length of the RC measured in number of words.

I should mention that cross-linguistic studies have mostly analyzed complexity of RCs resulting from movement, locality, intervention and feature similarity as well as word order expectation (see, for example, Durrleman et al. 2016, Rizzi 2013, Levy et al. 2013). Because of the difficulty in measuring complexity in written texts whose verification needs processing tests performed on speakers, this study investigates the effect of RC length on the Swahili *amba* and tensed RC variation. It is assumed that longer sentences have more phrases, clauses and optional adjuncts that require subordination and coordination (cf. Hemforth et al. 2015, who investigate the effect of length on high attachment (first noun modification) and low attachment (second noun modification) in the processing of relative clauses with two possible antecedents in English, Spanish and French).

3. Whether the grammatical role (subject/object) of the head noun in the matrix clause and information status of the head NP, specifically topic, impacts on the type of RC used.

There are some parameters that may help in identifying the topic in an utterance. These include grammatical role, pronominalization, and linear order of discourse entities. I posit that the notion of topic, defined as what the proposition is about (Gundel 1985, Lambrecht 1994), play a role on the choice of one form of the RCs over the other. Subjects are mostly topics while objects may be part of a comment in a topic-comment utterance structure. Based on qualitative and quantitative analysis of the dataset, I claim that the *amba* RC is preferred if an NP in a previous utterance is the topic of the RC in a shift transition while the tensed relative is preferred if the most salient NP (topic) of the matrix clause is also the topic of the RC.

The rest of the paper is organized as follows. In §2 I discuss previous work that explains the variation of the Swahili *amba* and tensed RCs. In §3 I briefly explain the methodology. I present the results of the study and discussion in §4 followed by the conclusion in §5.

2 Previous studies on the Swahili RC variation

The Swahili RC variation has mostly been attributed to morphosyntactic restrictions on the tensed RM. While the *amba* RM may be used with all the tense, aspect and modality (TAM) markers, the tensed RC may only be used with the past tense marker *li,* present tense marker *na*, the future tense marker *taka*, and the *si* negation marker (see Keach 1985 for a detailed description of TAM markers that require the use of *amba*). Due to the verb status of the tense markers and *amba* from a diachronic perspective, it has been argued that the two Swahili RCs are structurally the same (Vitale 1981, Keach 1985, Demuth & Harford 1999, Ngonyani 2001, 2006). A subject postposing rule in object RCs have been used as evidence to argue for the same syntactic position of the Swahili RM. Example (6) and (7) show the *amba* and tensed object RC respectively.

(6) Kikosi [amba-cho Kisaka a-li-ki-ongoza].
   squad amba-7.ORM 1.Kisaka 1.SM-PST-7.OM-lead
   'The squad which Kisaka led.'

(7) Kikosi [a-li-cho-ki-ongoza Kisaka].
   7.squad 1.SM-PST-7.ORM-7.OM-lead 1.Kisaka
   'The squad which Kisaka led.'

The prefix *a* attached to the verb *ongoza* 'lead' corresponding to the noun class 1 NP indicates that *kisaka* is the subject of the RC while the RM *cho* corresponding

to the noun class 7 NP *kikosi* indicates that the object NP marked by the prefix *ki* before the root of the verb is the relativized NP. In (6) the subject, *kisaka,* is preverbal but postverbal in (7). Notice that the head noun *kikosi* 'squad' occurs before *amba*/tense (*li*). Keach (1985) argues that the tensed RM and the *amba* RM are both independent words (verbs) and that nothing can intervene between the head noun and *amba*/tense, hence, the postverbal position of the subject in the tensed object RC in (7). However, the structural similarities/differences may not explain why one of the RCs may be chosen instead of the other in discourse texts.

Russell (1992: 125–126) argues that *amba* is used in object RCs to disambiguate the subject and object NPs if they both belong to the same noun class as illustrated in (8) and (9).

(8) Mwivi [a-li-ye-mw-ona   mtoto].
    1.thief 1.SM-PST-1.RM-1.OM-see 1.child
    'The thief who saw the child/The thief whom the child saw.'

(9) Mwivi [amba-ye   mtoto a-li-mw-ona].
    1.thief amba-1.ORM 1.child 1.SM-PST-1.OM-see
    'The thief whom the child saw.'

The NPs *mwivi* 'thief' and *mtoto* 'child' agree in number and class with the verb *a-li-mu-ona* and can therefore be both interpreted as subject/object in (8) since Swahili allows both the SVO and OVS word order. The use of *amba* licenses the occurrence of the subject in its canonical preverbal position as seen in (9) and therefore enhances the interpretation of *mtoto* 'child' as the subject of the RC. However, the disambiguation justification for the choice of *amba* is limited in its application since cross-reference can establish the subject and object grammatical roles in situations where the two belong to different noun classes as seen in (10).

(10) Mwivi wa-li-ye-mu-ona   watoto.
     1.thief 2.SM-PST-1.ORM-1.OM-see 2.children
     'The thief who the children saw.'

In (10), the subject of the RC is *watoto* 'children' because the subject marker *wa* on the verb agrees with the noun class 2 NP *watoto* 'children'. Furthermore, as was mentioned by an anonymous reviewer, although *amba* is used in (11) the subject of the sentence remains ambiguous if the NP *mtoto* 'child' occurs after the verb.

(11)   Mwivi [amba-ye   a-li-mw-ona mtoto].
        1.thief amba-1.ᴏʀᴍ 1.ꜱᴍ-ᴘꜱᴛ-see 1.child
        'The thief whom the child saw.'

Considering examples such as (10) and (11), the choice of the *amba* relative may not be necessarily to disambiguate the subject/object grammatical roles. Furthermore, the grammatical role (subject/object) and function (agent/theme) of discourse entities belonging to the same/different noun classes can be identified if adequate contextual background is available. In texts, discourse entities are linked via referring expression and therefore parameters such as linear order of utterances, subjecthood and pronominalization can help identify the intended referents as well as their grammatical role and information statuses (cf. Grosz et al. 1995 and Grosz & Sidner's 1998 Centering Theory, which attempts to determine topic in discourse texts).

In Swahili, it has been claimed that the *amba* RC is preferred if the RC in question is complex as illustrated in (12) and (13).

(12)   Masanduku [amba-yo   sisi watu   wa-wili   au wa-tatu
        6.boxes   amba-6.ᴏʀᴍ us 2.people 2.ᴀɢʀ-two or 2.ᴀɢʀ-three
        ha-tu-wez-i]₁   [ku-ya-enua]₂.
        ɴᴇɢ-1ᴘʟ-able-ꜱʙᴊᴠ ɪɴꜰ-6.ᴏᴍ-lift
        'Boxes which two or three of us cannot lift.'

(13)   Masanduku [tu-si-yo-weza   sisi watu   wa-wili   au watatu]
        6.boxes   1.ᴘʟ-ɴᴇɢ-6.ᴏʀᴍ-able us   2.people 2.ᴀɢʀ-two or 2.ᴀɢʀ-three
        ku-ya-enua]₂.
        ɪɴꜰ-6.ᴏᴍ-lift
        'Boxes which two or three of us cannot lift.'   (Ashton 1944: 310)

Following Keenan & Comrie's (1977) Accessibility Hierarchy of relativization, it has been claimed that the *amba* RC in (12) is preferred instead of the tensed RC in (13) since the relativized NP is the object (which is higher in the Accessibility Hierarchy than the subject). Based on the dataset used in this study, I argue that the choice of the *amba* RM in (12) is influenced by the length of the RC rather than the position of the head noun in the Accessibility Hierarchy. Notice that the RC in (12) contains two clauses with 8 words. The *amba* relative clause may be used by the speaker to "compensate for (perceived) difficulties in producing or parsing" long relative clauses, whether subject or object RCs (Green 1996: 133).

In these syntactic studies on Swahili RC, the structural differences/similarities are the focus and their usage by speakers is regarded as a case of free variation.

However, a combination of pragmatic factors such as preference for the tensed RC in restrictive RCs, length of the RCs, and information structure considerations may better explain the choice of one of the RC variants in natural language.

3 Methodology

The source of data in this study was the Helsinki Corpus of Swahili (HCS) which has 25 million words. The corpus contains Swahili newspaper articles as well as excerpts of literary texts, education and science material written from the mid-20th century to 2015. In the HCS, concordance searches are done via an inbuilt software, namely *Korp. Korp* displays a concordance list of the query as well as the immediate context of the search expression. Every word in the Helsinki corpus contains information related to its parts of speech (for example; verb, noun, adjective), morphological description (for example, relative marker, reciprocal, tense), syntactic function (for example; subject, main verb, object), and gloss in English. The tensed RCs were displayed using an extended search that instructed *Korp* to look for words whose parts of speech was V (verb) and its morphology included REL (relative marker). The *amba* RCs were displayed using a search that queried for all occurrences with *amba* as the base form. Only TAM markers that allow the occurrence of both the *amba* and tensed RCs were selected. Further, the RCs were selected such that they represent the diverse sources in the corpus which include novels by different authors, newspapers, and academic articles on literary works. In total, 440 RCs were selected from the HCS; 285 were tensed RCs and 155 were *amba* RCs.

Recall that *Korp* displays the target utterances in their immediate context. The utterances that occur within the context of the target RC provide contextual clues that help in the pragmatic analysis of the Swahili RC variation. Each of the 440 RCs was coded as restrictive or non-restrictive. As explained earlier, a restrictive relative gives specific information that delimit a referent while a non-restrictive clause gives more information (non-identifying) about a referent. The number of words within each RC was then counted to measure its length. After identifying the discourse entities in the target utterance and those occurring before and after it the RC was coded for relativization (subject/object), the grammatical role of the head noun in relation to the matrix clause (subjet/object/PP compliment) and information status of the head NP (topic/non-topic).

4 Results and discussion

The study aimed at investigating pragmatic reasons for the variation of Swahili *amba* and tensed RCs. I discuss the results in the following order. In §4.1, I discuss the role of restrictive/non-restrictive on the *amba* RC and tensed RC variation. In §4.2, I discuss the effect of the RC length on the choice of the *amba*/tensed RC. In §4.3 I show that information structure, hitherto unexplored, may also impact on the Swahili RC variation.

4.1 Variation due to restrictive/non-restrictive use of RCs

If the variation of the *amba* and tensed RCs is due to the restrictive and non-restive use of the RCs, I expect the dataset used in the study to have a higher frequency of *amba* non-restrictive RCs than the tensed non-restrictive RCs. On the other hand, I expect a higher frequency of tensed restrictive RCs than *amba* restrictive RCs. Of the 155 *amba* RCs, 32 were restrictive and 123 were non-restrictive. As for the 285 tensed RCs, 245 were restrictive and 40 were non-restrictive. Figure 1 summarizes these results.
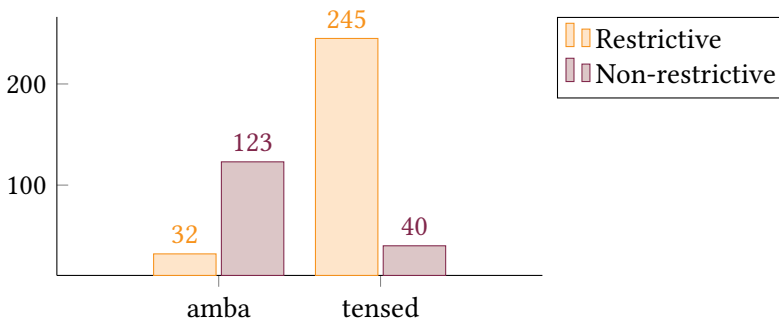

Figure 1: *amba* and tensed restrictive and non-restrictive clauses

A chi-squared test shows that the frequency differences between the restrictive and non-restrictive clauses for the *amba* and tensed RCs is significant $\chi^2(1, N = 440) = 180.89, p < 0.001$. The implication is that the tensed RC is used more frequently as restrictive while the *amba* RC is used more frequently as non-restrictive. I present an example of a tensed restrictive RC in (14).

(14) Bwana yule [a-na-ye-sinzia].
     1.man   that 1.SM-PRS-1.SRM-doze
     'That man (over there) who is dozing.'

In (14), the speaker points at a specific man who was dozing amongst other men. The RC gives information that help in delimiting the referent of the referring expression *bwana yule* 'that man'. Although the tensed RC was mostly used restrictively, there are a few instances of tensed non-restrictive RCs illustrated in (15).

(15) A-li-m-rud-i-a Adili [a-li-ye-kuwa a-na-pata
     1.SM-PST-1.OM-return-APPL-FV 1.Adili 1.SM-PST-1.SRM-AUX 1.SM-PRS-get
     fahamu].
     9.consciousness
     'He got back to Adili, who was regaining his consciousness.'

The RC in (15) does not delimit the referent but gives more information about the topic of the RC, *Adili*. A close analysis of the non-restrictive tensed RCs shows that they mostly occur if the referent of the head noun is a proper noun as is the case in (15). There are also cases when a non-restrictive tensed RC gives more information that establishes the identity of a proper noun, as seen in (16).

(16) Krapf [a-li-ye-li-tumikia dhehebu la C.M.S].
     Krapf 1.SM-PST-1.SRM-5.OM-serve 5.denomination of C.M.S.]
     'Krapf, who served the C.M.S Christian denomination.'

In (16), the head noun *Krapf* is the subject of the RC. The RC presents more specific information that may assist the hearer identify the intended referent.

While the dataset shows that the tensed RC mostly gives information that delimits the referent in restrictive RCs or gives identifying information in non-restrictive usage, the *amba* RC adds new non-identifying information predicated on the head noun as seen in (17).

(17) A-li-kuwa a-me-va-li-a rubega nyeupe [amba-yo
     1.SM-PST-AUX 1.SM-PFV-wear-APPL-FV 9.cloak white amba-9.SRM
     i-li-anza ku-poteza weupe wake].
     9.SM-PST-start INF-loose whiteness its
     'He was wearing a white cloak that had started to lose its whiteness.'

In (17), the RC gives more information about the head noun, *rubega nyeupe* 'white cloak', whose whiteness had faded. The RC in (17) may be analyzed as an utterance with a topic-comment structure rather than an NP in a coreferential relationship with its antecedent. Although the *amba* RC is mostly non-restrictive, there are instances when the *amba* RC may be used restrictively as seen in (18).

(18)   Mambo muhimu   [[amba-yo   ya-na-fanya]$_1$ [shairi
       6.things important amba-6.SRM 6.SM-PRS-make 6.poem
       li-it-w-e   la Ki-swahili]$_2$].
       6.SM-call-PASS-SBJV of 7-Swahili
       'Important things that make a poem be called a Swahili poem.'

In (18) the *amba* RC is restrictive because the head noun *mambo muhimu* 'important things' are defined by the RC as a set of characteristics that make a poem be called a Swahili poem. Notice that compared to the tensed restrictive clause presented in (14), the *amba* restrictive clause in (18) is longer. The tensed RC is one-word long (one clause) while the *amba* RC is six-words long (two clauses) and both are restrictive and subject RCs. I present the results of the effect of length on Swahili RC variation in §4.2.

4.2 Length of the relative clause

In this section, I discuss the effect of the RC length measured in number of words on Swahili *amba* RC and tensed RC variation. I claim that the *amba* RC is more frequently used if the length of the RC is long, while the tensed relative is used more frequently in short RCs as illustrated in (19) and (20).

(19)   a. Mwanamke mjane   amba-ye   a-li-kuwa   jirani   yao.
           1.woman   1.widow amba-1.SRM 1.SM-PST-be 1.neighbor their
           'A (woman) widow who was their neighbor.'

       b. Mwanamke mjane   a-li-ye-kuwa   jirani   yao.
           1.woman   1.widow 1.SM-PST-1.SRM-be 1.neighbor their
           'A (woman) widow who was their neighbor.'

(20)   a. Mwanamke mjane   amba-ye   a-li-kuwa   jirani   yao hapo
           1.woman   1.widow amba-1.SRM 1.SM-PST-be 1.neighbor their long
           zamani.
           ago
           'A (woman) widow who was their neighbor long ago.'

       b. Mwanamke mjane   a-li-ye-kuwa   jirani   yao hapo zamani.
           1.woman   1.widow 1.SM-PST-1.SRM-be 1.neighbor their long ago
           'A (woman) widow who was their neighbor long ago.'

The RCs in (19a–b) are shorter than the RCs in (20a–b) because of the addition of the adverbial phrase *hapo zamani* in the later. Regardless of the grammatical

role of the head noun, the dataset indicates that the *amba* RC was mostly used in cases where the length of the RC in question was long as seen in (20). I claim that the *amba* relative is used more frequently in such cases due to its simplifying effect in parsing or producing long RCs.

Table 1 presents the mean and standard deviation (in parenthesis) of the length of the *amba* RCs and tensed RCs (restrictive and non-restrictive).

Table 1: Mean and sd of RC length in restrictive and non-restrictive clauses

	Restrictive		Non-restrictive	
amba	6.34	(3.25)	7.21	(3.71)
tensed	3.09	(2.34)	4.25	(2.61)

The mean length of the *amba* RCs is higher (about 7 words) than that of the tensed RCs (about 4 words). A non-repeated measures ANOVA with RC length as the independent variable, and RM (*amba*/tensed) and relative type (restrictive/non-restrictive) as the dependent variables reveal that there is a significant main effect of RC length on RM, $(1, 440) = 197.49$, $p < 0.001$, and relative type $F(1, 440) = 11.11$, $p < 0.001$. These results indicate that the *amba* RC is preferred in long RCs while the tensed RC is preferred in short RCs.

In addition to its simplifying effect in parsing and production of long RCs, the preference for *amba* in such situations is also due to its ability to allow for more types of phrases to occur in different orders as illustrated by the adverbial phrase (AP) *kwa makusudi* 'deliberately' in (21–24).

(21)   Wa-po   wananchi [amba-o   kwa makusudi wa-na-ziba
        2.sm-16.loc 2.citizens [amba-2.srm with 6.deliberate 2.sm-prs-block
        mifereji].
        4.taps]
        'There are some citizens who deliberately block taps.'

(22)   * Wa-po   wananchi kwa makusudi   [wa-na-o-ziba   mifereji].
        2.sm-16.loc 2.citizens with 6.deliberate [2.sm-prs-2.srm-block 4.taps]

(23)   Wa-po   wananchi [amba-o   wa-na-ziba   mifereji kwa
        2.sm-16.loc 2.citizens [amba-2.srm 2.sm-prs-block 4.taps   with
        makusudi].
        deliberate]

(24)   Wa-po   wananchi [wa-na-o-ziba   mifereji kwa makusudi].
       2.SM-16.LOC 2.citizens [2.SM-PRS-2.SRM-block 4.taps   with deliberate]

In (21), the *amba* RC must be used because of the fronted AP *kwa makusudi* 'deliberately'. The use of the tensed RC with the AP intervening between the head noun *wananchi* 'citizens' and the RM in *wa-na-o-ziba* renders (22) ungrammatical. Both the *amba* RC and the tensed RC are grammatical if the adverbial phrase is post-verbal as in (23) and (24). The different word order with the AP in sentence final position may however change the intended effect of the utterance as stipulated by information structure considerations discussed in §4.3.

4.3 Effect of grammatical role and topic

I claim that in addition to the restrictiveness and RC length, the use of the *amba*/ tensed RC may also be influenced by information structure considerations. Information structure, a term originally introduced by Halliday (1967), deals with formal properties of semantically equivalent but pragmatically divergent propositions in their textual environment (Lambrecht 1994). The notion of topic, defined as what the sentence is about, has a role in the choice of Swahili RC constructions. Under information structure, utterances are organized into two parts. In the unmarked form, the first part, the topic, is given or presupposed information while the second part, comment, the predicate and its internal argument(s) and adjunct(s) is the new or informative part of the utterance (Prince 1981, 1992, Gundel & Fretheim 2006). To investigate the role of grammatical role and topic in the variation of the tensed and *amba* RCs, I coded the relativization of the head noun (subject/object/prepositional phrase (PP)). Based on its grammatical role and other contextual clues including pronominalization and referential chain (previous or later mention), the head noun was coded as topic/non-topic. This was possible because each utterance has exactly one topic which is the most salient member of the referents realized as pronoun, explicit referring expression, zero or "inferred" (as used by Prince 1981). Salience "defines the degree of relative prominence of a unit of information, at a specific point in time, in comparison to other units of information" (Chiarcos et al. 2011). The concept of topic as the most salient entity of an utterance is illustrated in (25).

(25)   a. Nyumba [a-li-yo-kaa   Beneti].
          9.house 1.SM-PST-9.ORM-stay Beneti
          'The house which Beneti stayed.'

    b. I-li-kuwa   ya Kiswahili kabisa,
       9.sm-pst-be of Swahili   typical
       'It was a typical Swahili house,'

    c. I-li-on-ek-an-a   kuukuu.
       9.sm-pst-see-stv-recp-fv old
       'it looked old.'

In (25), there are three clauses labeled as a, b and c. In (25a), the subject of the RC is the postposed NP *Beneti* and the relativized NP *nyumba* is the object. Although it is possible to use the *amba* RM in (25a), I argue that the tensed RC is chosen so that the subject NP *Beneti* is postposed to place *nyumba* 'house' as the sole preverbal NP to make it easier for the hearer to parse *nyumba* as the topic of the matrix clause. The head noun *nyumba* 'house' is the subject of the utterance in (25) and therefore occurs in (25b–c) as a pronominal subject prefix attached to the verb. The most salient information unit of the matrix clause is therefore the house because of its occurrence as the sole NP in preverbal position and its pronominalization.

A restrictive clause is analyzed as a modifier of the head noun and therefore an element within a noun phrase structure of a subject, object or PP complement. On the other hand, a non-restrictive clause is analyzed as an independent clause with its own locus for topic update. The head noun may be the topic/non-topic of the RC in question. While (25) shows a head noun, *nyumba* 'house', that is the topic of the matrix clause, (26) shows a head noun that is non-topic.

(26)   Adili a-li-vunja   amri   [a-li-yo-p-ew-a].
       1.Adili 1.sm-pst-break 9.command 1.sm-pst-9.orm-give-pass-fv
       'Adili broke the command he was given.'

In (26), the head noun *amri* is the object of the matrix clause and together with the RC is an internal argument of the verb *vunja* 'break'. The topic of the matrix clause is the subject NP *Adili* in the left periphery. The occurrence of the NP *Adili* as a pronominal subject marker prefix within the RC provides further evidence for the topic status of the NP *Adili.* In instances with no topic shift such as (25) and (26), the tensed RC is more frequently used regardless of the grammatical role of the head noun. In both instances, the head noun is the object of the RC, an RC that is predicted to be complex by the Accessibility Hierarchy (Keenan & Comrie 1977) and therefore the *amba* RC more appropriate due to its simplification effects (Ashton 1944, Schadeberg 1989). Notice that the length of the tensed RCs in (25) and (26) are short (one word long with no subordination or coordination). I argue

that the use of the tensed clause instead of the *amba* RC is due to the length of the RCs as explained in §4.2 and the information status of the head noun.

Table 2 shows the results of the effect of grammatical role and topic on the frequencies of *amba* and tensed restrictive/non-restrictive RCs. Row one, for example, shows the frequencies of the *amba* RCs and tensed RCs (restrictive and non-restrictive) when the head noun (HN) is the subject (topic) of the matrix clause and is also the topic of the RC in question.

Table 2: Effect of grammatical role and topic on Swahili RC variation

	amba		tensed		Total
	restr.	non-restr.	restr.	non-restr.	
HN = Subj, topic	7	21	62	8	98
HN = Subj, non-topic	1	0	4	1	6
HN = Obj, topic	10	55	21	15	101
HN = Obj, non-topic	3	24	94	5	126
HN = PP, topic	5	18	9	9	41
HN = PP, non-topic	6	5	55	2	68

The frequency of head NPs that were the topic of the matrix clause and were continued as the topic of the tensed RCs was higher (70) than that of head nouns that were topics in the matrix clause as well as the *amba* RCs (28). The frequency difference is significant, $\chi^2(1, N = 98) = 18, p < 0.001$. Example (27) shows a tensed RC whose head noun is the topic of the matrix clause as well as the RC.

(27) Waandishi [wa-li-o-fuata]   wa-me-ach-an-a   na m-tindo
      2.writers   2.SM-PST-2.SRM-follow 2.SM-PFV-leave-RECP-FV with 3-style
      huu.
      this
      'The writers who followed have abandoned this style (of writing).'

In (27), the head noun *waandishi* 'writers' is the subject (topic) of the matrix clause and is continued as the topic of the RC via the subject marker *wa*. RCs that delimit the referent of a topic NP in a continue transition are mostly tensed. However, the *amba* RC may be used in long RCs even if the head noun is subject of the matrix clause and topic of the RC in question as discussed in §4.2.

There was a total of 101 RCs (65 *amba* and 36 tensed) whose head nouns were objects of the matrix clause but topic of the RC in question. On the other hand,

there were 126 RCs (27 *amba* and 99 tensed) whose head nouns was object of the matrix but were non-topics within the RC utterances. The *amba* RC was used to enhance a shift transition in the event that the object NP became the topic of the RC as illustrated in (28).

(28) [Waandamanaji wa-li-kaidi   amri   hiyo]₁ [na
      2.demonstrators 2.SM-PST-disobey 9.command that   and
      ku-wa-rush-i-a   mawe   polisi]₂   [amba-o
      INF-OM-throw-APPL-FV 6.stones 2.policemen amba-2.SRM
      wa-li-jibu   mapigo]₃ [kwa ku-fyatua risasi   za moto]₄.
      2.SM-PST-respond 6.beatings by   INF-shoot 10.bullets of 3.fire
      'The demonstrators disobeyed the command and threw stones to the
      policemen who responded to the beatings (of stones) by shooting live
      bullets.'

In (28), there are four clauses numbered 1–4. The topic of clause 1 and 2 is *waandamanaji* 'demonstrators' while the topic of clause 3 and 4, the *amba* RC, is *polisi* 'policemen', the object of the matrix clause. Since the NP *polisi* 'police' is mentioned for the first time as the object of the matrix clause in clause 2, then functions as the subject of the following clause, the *amba* RC, then it becomes the most salient NP in that clause due to its subject role and givenness (previously mentioned). The *amba* clause is preferred in a shift transition because of the topic status of the previously mentioned object NP to assist in parsing the RC as a separate utterance with a topic-comment structure. Although the frequency of the *amba* RCs used in shift transition is significantly higher than that of the tensed RCs, the tensed RC may also be used as shown in (29).

(29) Bali Rehema   a-li-po-nyanyua   macho ku-m-tazama Bikiza
      but 1.Rehema 1.SM-PST-when-raise 6.eyes INF-2.OM-look 1.Bikiza
      [a-li-ye-leta   taarifa   ya kifo]...
      1.SM-PST-1.SRM-bring 9.message of 7.death
      'But when *Rehema* raised her eyes to look at Bikiza who brought the
      meassage of death...'

In (29), *Bikiza* is the object of the matrix clause but the topic of the tensed RC. I mentioned earlier that the tensed non-restrictive RC occur in contexts where the head noun is a proper noun. This may explain the choice of the tensed RC here, but I should reiterate that the pragmatic explanations in this study are not absolute. Of course, they are statistical tendencies that cannot be attributed to chance.

Mohamed Mwamzandi

41 head nouns occurring as PP complements within the subject/object NP of
the matrix clause were RC topics while 68 were non-topics. The frequency dif-
ference between the *amba* and tensed RCs when the head NP was topic was in-
significant, $p > 0.05$. However, the frequency difference in non-topic head nouns
was significant ($\chi^2(1, N = 77) = 31, p < 0.001$). In (30), the *amba* RC is used in
shift transitions, that is, the head noun is not the topic of the matrix clause but
is the topic of the RC.

(30) A-li-anza   ku-safiri ili   ku-epuka ghadhabu ya ndugu   yake
      1.SM-PST-begin INF-travel so.that INF-avoid 9.anger   of 1.sibling his
      amba-ye   a-li-u-kosa   u-rithi   huo.
      amba-1.SRM 1.SM-PST-11OM-miss 11-inheritance that
      'He started traveling to avoid the anger of his sibling who had missed the
      inheritance.'

In (30), the head noun *ndugu yake* 'his sibling' is the complement of the prepo-
sition *ya* 'of'. The subject of the matrix clause is a discourse entity occurring
as subject pronominal NP *a* (The co-referential NP is not overt in the example.)
while the subject of the *amba* RC is *ndugu yake* 'his sibling' occurring as a the
SRM *ye* and pronominal NP *a* within the RC. The use of *amba* in (30) enhances
the parsing of the RC as an utterance with its own locus for update rather than
a delimiting phrase of the object NP. A close analysis reveals that tensed RCs
whose head is topic of the RC but a PP complement of the matrix subject/object
NP were mostly predicated by intransitive or passivized verbs as illustrated in
(31) and (32).

(31) A-li-ji-ona   katika nyumba kubwa sana i-li-yo-kuwa
      1.SM-PST-REFL-see in   9.house 5.big   very 9.SM-PST-9.SRM-be
      n-zuri   ajabu.
      9-beautiful surprisingly
      'He found himself in a very big house that was surprisingly beautiful.'

(32) Katika hukumu i-li-yo-to-lew-a   jana   na Jaji
      in   9.ruling 9.SM-PST-9.SRM-give-PASS-FV yesterday by 1.judge
      Bernard Luand.
      Bernard Luand
      'In a ruling that was given yesterday by Judge Bernard Luand.'

In (31) the subject/topic of the tensed RC is the NP *nyumba kubwa sana* 'a very big house' which is a complement of the preposition *katika* 'in'. In (32) the head noun is the NP *hukumu* also occurring after the preposition *katika* 'in'. As for the head nouns which were PP complements of subject/object NPs of the matrix clause, the dataset generally indicates that the tensed RC in non-topic head nouns is the most frequently used.

5 Conclusion

A combination of factors including restrictive/nonrestrictive use of the RC, length of the relative, grammatical role of the head noun and information structure may explain the choice of one of the Swahili RC variants in cases that allow both the *amba* and tensed RC, specifically, when the past, present and future tense markers are used. The dataset suggests that all things being equal, the tensed RM is mostly used if the RC is restrictive while the *amba* RM is mostly used if the RC is non-restrictive. Further, the tensed RC can mostly be analyzed as a modifier of the head noun. On the other hand, the *amba* RC presents new information predicated on the relativized NP. The dataset also shows that the mean length of the *amba* RC is significantly higher than that of the tensed RC. The implication is that the *amba* RC contains more discourse entities and words in subordinated and coordinated structures. This calls for further processing and reading experiments to find out whether the *amba* RM is used in RCs to compensate for difficulty in production and parsing of longer utterances. Another factor that impacts on the choice of Swahili RCs is the grammatical role and information status of the head noun and other discourse elements within the RC. The *amba* RC is more frequently used in shift transition while the tensed relative delimits references of topic/non-topic discourse entities in continue transition. It is also possible that the tensed RC is preferred if the predicate is intransitive in shift transition, but this observation needs further research.

Source

Helsinki Corpus of Swahili 2.0 (HCS 2.0). 2014-05-09. User support at CSC – IT Center for Science Ltd. The Language Bank of Finland (distributor). Etsin research data finder, 2018. http://urn.fi/urn:nbn:fi:lb-2014032624

Mohamed Mwamzandi

Abbreviations

1, 2, 3, ...	Noun classes	PFV	Perfective
AGR	Agreement	PROG	Progressive
APPL	Applicative	PRS	Present tense
AUX	Auxiliary	PST	Past tense
FUT	Future tense	RECP	Reciprocal
FV	Final vowel	REFL	Reflexive
INF	Infinitive	SBJV	Subjunctive
OM	Object Marker	SM	Subject marker
ORM	Object relative marker	SRM	Subject relative Mark
PASS	Passive	STV	Stative

References

Andrews, Avery. 2007. The major functions of noun phrase. In Timothy Shopen (ed.), *Language typology and syntactic description*, 132–223. New York, NY: Cambridge University Press.

Ashton, Ethel O. 1944. *Swahili grammar (including intonation)*. London: Longmans, Green & Co Ltd.

Chiarcos, Christian, Claus Berry & Michael Grabski. 2011. *Salience: Multidisciplinary perspectives on its function in discourse*. Berlin/New York: De Gruyter Mouton.

Comrie, Bernard. 1989. *Language universals and linguistics typology*. Oxford: Basil Blackwell Publisher Limited.

Demuth, Kathrene & Carolyne Harford. 1999. Verb raising and subject inversion in Bantu relatives. *Journal of African Languages and Linguistics* 20(1). 41–61. DOI: 10.1515/jall.1999.20.1.41.

Durrleman, Stephanie, Theodoros Marinis & Julie Franck. 2016. Syntactic complexity in the comprehension of *wh*-questions and relative clauses in typical language development and autism. *Applied Psycholinguistics* 37(6). DOI: 10.1017/S0142716416000059.

Green, Georgia M. 1996. *Pragmatics and natural language understanding*. New Jersey: Lawrence Erlbaum Associates Inc.

Grosz, Barbara J. & Candace L. Sidner. 1998. Lost intuitions and forgotten intentions. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering theory in discourse*. New York, NY: Oxford University Press.

Grosz, Barbara J., Scott Weinstein & Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.

Gundel, Jeanette K. 1985. "Shared knowledge" and topicality. *Journal of Pragmatics* 9(1). DOI: 10.1016/0378-2166(85)90049-9.

Gundel, Jeanette K. & Thorstein Fretheim. 2006. Topic and focus. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 175–196. Malden, MA: Blackwell Publishing.

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part II. *Journal of Linguistics* 3(2). 199–244.

Hemforth, Barbara, Susana Fernandez, Charles Clifton Jr., Lyn Frazier, Lars Konieczny & Michael Walter. 2015. Relative clause attachment in German, English, Spanish and French. 166. 43–64. DOI: 10.1016/j.lingua.2015.08.010.

Huang, Yan. 2007. *Pragmatics*. New York, NY: Oxford University Press.

Keach, Camillia. 1985. *The syntax and interpretation of the relative clause construction in Swahili*. New York, NY: Garland Publishing Inc.

Keenan, Edward. 1985. Passives in the world's languages. In Timothy Shopen (ed.), *Language typology and syntactic description*, 241–281. New York, NY: Cambridge University Press.

Keenan, Edward & Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99. https://www.jstor.org/stable/4177973.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Levy, Roger, Evelina Fedorenko & Edward Gibson. 2013. The syntactic complexity of Russian relative clauses. *Journal of Memory and Language* 69(4). 461–495.

Ngonyani, Deo. 2001. Evidence for head raising in Kiswahili relative clauses. *Studies in African Linguistics* 30(1). 59–73. https://journals.flvc.org/sal/article/view/107362.

Ngonyani, Deo. 2006. Attract F and verbal morphology in Kiswahili. *The Linguistic Review* 23(1). 37–68. DOI: 10.1515/TLR.2006.002.

Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–256. New York, NY: Academic Press.

Prince, Ellen. 1992. ZPG letter: Subjects, definiteness, and information-status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse description: Diverse analyses of a fund raising text*, 295–326. Philadelphia, PA/Amsterdam: John Benjamins.

Rizzi, Luigi. 2013. Locality. *Lingua* 130. 169–186. DOI: 10.1016/j.lingua.2012.12.002.

Russell, Joan. 1992. From reanalysis to convergence: Swahili-amba. *York Papers in Linguistics* (16). 121–138.

Schadeberg, Thilo C. 1989. The three relative constructions in Swahili (kisanifu). In Marie-François Rombi (ed.), *Le swahili et ses limites: Ambiguïte des notions reçues*, 33–40. Paris: Editions Recherche sur les Civilisations.

Vitale, Anthony. 1981. *Swahili syntax*. Cinnaminson, NJ: Foris Publications.