

Chapter 2. Morphosyntax of Swahili

Introduction

A meaningful analysis of child language is impossible without a clear understanding of the adult language. This chapter is divided into two sections. In the first section, I will start out by discussing some social and cultural aspects of Swahili, followed by the morphological characteristics of Swahili: its noun class system, the agreement system, the affixes which make up the verbal complex, etc. Since this dissertation investigates the acquisition of Swahili verbal morphology, this section will focus more on the descriptions of verbal morphology than nominal morphology, as it will be necessary to draw on these descriptions in later chapters. In the second section of this chapter I will present my theoretical assumptions, sketch out a syntactic analysis of Swahili functional structure, and discuss how this analysis fits in with some current debates in the Bantu literature, e.g., whether subject agreement marking is actually agreement or a pronominal clitic. I will then discuss the omission of subject agreement in adult Swahili and propose an analysis of null subjects in these clauses. This section will be particularly relevant in chapter 5 where I discuss Subject Agreement omission in child Swahili. The purpose of this chapter is not to provide an exhaustive analysis of Swahili morpho-syntax, but rather to provide a reader who has little or no knowledge of the Bantu languages with enough information to adequately understand the subsequent acquisition chapters.

Swahili has a complex and controversial status in Eastern Africa today. There are currently approximately 50 million speakers of Swahili (Hinnebusch, 1979), of which 2 million are native speakers (the remainder

being proficient second language speakers). This reflects the history of Swahili, as it was used as a trading language for those who traveled from the ports of Mombasa, Dar es Salaam and Zanzibar into the interior of the African continent. These two million speakers are primarily inhabitants of the coastal regions of Kenya and Tanzania, including Zanzibar. However, speakers in these areas speak slightly different dialects of Swahili. Swahili found in and around Mombasa is called *Mvita*, and that spoken in Zanzibar and the surrounding coastal mainland is called *Unguja*. Modern Standard Swahili, or Kiswahili Sanifu, is based on *Unguja*.

However, Kiswahili Sanifu and the other ‘standard’ dialects of Swahili are not the only forms of Swahili spoken in the region. Indeed, they are less widespread than the more colloquial, less socially accepted dialects of inland Kenya and Tanzania. Kiswahili Sanifu is used primarily by the mass media, in school textbooks and exams, and by the governments of Tanzania and Kenya. The other dialects are used in day-to-day conversation and communication between neighboring tribal and ethnic groups. These dialects are the true language of communication.

Nairobi, the city in which the participants in this study were being raised, is an extremely socially, ethnically and tribally diverse city. Swahili spoken in Nairobi is the product of this diverse environment, and differs significantly from Kiswahili Sanifu. Swahili spoken in Nairobi ranges in a continuum from dialects that are almost standard to dialects that are much closer to so-called ‘pidgin’ Swahili (see Duran, 1975). In this study, when I refer to Nairobi Swahili, I am referring to the dialect of the subjects in this

study. I make no claim that Swahili spoken in Nairobi constitutes a single, unified dialect.¹

The so-called ‘slum’ areas of Nairobi are the true neighborhoods of Nairobi, as the majority of the population lives in one of these many low-income neighborhoods that surround the city. The dialect of Swahili in this study is the product of these eclectic residential areas, which are characterized by communal, close-quarters living. The children in this study reside in outlying neighborhoods in Nairobi (two in Majengo, one in Riruta, and one in Komarock), and so the only language they were exposed to was this dialect of Swahili. They hardly ever had the chance to watch television, and of course were too young to read newspapers. Furthermore, it is not a culturally common practice for adults to read to children, and so these children were rarely exposed to standard Swahili.

There are clear linguistic differences between Nairobi Swahili (i.e., the particular dialect of Swahili spoken by these subjects) and Kiswahili Sanifu. Kiswahili Sanifu has a richer agreement system and a richer noun class system than Nairobi Swahili (9 in Nairobi Swahili, as opposed to the 15 traditionally ascribed to Kiswahili Sanifu – see table 2.1 below). In addition, because of the dynamic social conditions in which Nairobi Swahili exists, there has been extensive borrowing from English, Kikuyu and Luo (the two other major African languages in Kenya). A comparative syntactic analysis of the two dialects is beyond the scope of this dissertation, and so I

¹ The dialects of the families in this study are not significantly different from each other (see appendix 3A). Where this dialect falls on the continuum is not crucial to our purposes here, but I believe it is closer to the non-standard end than to Kiswahili Sanifu. Thanks to Thomas Hinnebusch for discussion on this point.

will limit myself to a description and analysis of Nairobi Swahili, indicating when possible some major differences between the two dialects of Swahili.

I will now discuss some basic descriptive facts, followed by a description of the noun class system. I will then describe the verbal complex, taking each morpheme in turn and describing its form and possible functions. In the second section of the chapter, I will present the syntax of Swahili.

2.1 Basic Facts

Swahili is an agglutinative language, with considerable prefixing and suffixing. The unmarked word order is S-V-O, as shown in example (1)² below. In (1), the subject (Juma) occurs preverbally and the object (Mariam) occurs postverbally. The verb is embedded in a verbal complex which consists of subject agreement (*a-*) on the left periphery, followed by tense (*-na-*), object agreement (syllabic *-m-*) and then the verb root itself (*pend-*). The verb is followed by (in this case) one suffix which indicates mood (in this case indicative *-a*). The subject can be optionally absent (shown in example 2), and the person and number features of the subject are recoverable from the rich subject verb agreement. The subject may occur in postverbal position (3), with an obligatory pause and lower intonation (so-called comma intonation). Furthermore, the object may also be dropped (4).

- (1)

| | | |
|---------------------|---|--------|
| Subject | Verbal Complex | Object |
| Juma | a - na - m - pend - a | Mariam |
| Juma | SA _{3s} -Pres- OA _{3s} - like - IND | Mariam |
| 'Juma likes Mariam' | | |
- (2)

| | |
|---|--------|
| A - na - m - pend - a | Mariam |
| SA _{3s} - Pres- OA _{3s} - like- IND | Mariam |
| 'He likes Mariam' | |
- (3)

| | |
|--|---------------------------|
| ni - na - m - pend - a | Mariam, mimi |
| SA _{1s} - Pres- OA _{3s} - like - IND | Mariam Spro _{1s} |
| 'I like Mariam' | |
- (4)

| | |
|--|--|
| a - na - m - pend - a | |
| SA _{3s} -Pres - OA _{3s} - like - IND | |
| 'He likes her' | |

While I have described Swahili as an S-V-O language, there is a considerable amount of material that intervenes between the subject and the verb root, and the object and the verb root. However, as (3) above shows, when the subject is moved, all elements of the verbal complex (including Subject Agreement) remain with the verb. Similarly, if the object is preposed, as in (5) below, all elements of the verbal complex remain within the verbal complex in their original positions, including the object agreement marker:

- (5)

| | |
|--------------------------|--|
| Mariam, Juma | a - na - m - pend - a |
| Mariam, Juma | SA _{3s} - Pres- OA _{3s} - like - IND |
| 'Mariam, Juma likes [t]' | |

These examples show that the verbal complex behaves as a unit in Swahili. This will be discussed in more detail in the second half of this chapter in regards to the syntax of Swahili.

² Refer to the list of abbreviations after the table of contents for a guide to the glosses. Subscripts indicate agreement features/class between the SA

2.2 General phonological characteristics

Consonant clusters within syllables in Nairobi Swahili are not attested. Most syllables are open syllables, with the exception of homorganic nasals (Ashton, 1947; Myachina, 1981). Almost all Swahili words end in a vowel. This includes loan words, which in the original language end in a consonant, to which Swahili adds a vowel, e.g., *kitab* (Arabic for 'book') → *kitabū*, *television* → *televisheni*.

Most words are bisyllabic or trisyllabic, with monosyllabic words being avoided (Park, 1995; Myachina, 1981; Maw & Kelly, 1975).

Brandon (1975) argues this is because of a rule of penultimate stress that is quite widely adhered to.

- (6)

| | | |
|----|-------------|--------------------|
| a. | J í ko | fireplace, kitchen |
| b. | W á tu | people |
| c. | Ch ú pa | bottle |
| d. | P í ka | cook (v.) |
| e. | Kit á nda | bed |
| f. | Sab ú ni | soap |
| g. | Ang ú ka | fall (v.) |
| h. | Ong é a | talk (v.) |
| i. | Tegem é a | Depend (v.) |
| j. | Tafadh á li | Please |

There are few exceptions to the penultimate stress rule (mostly within the realm of loan words e.g., *lázima*, from Arabic, meaning

and the subject, or the OA and the object.

‘obligation’³), but this is a strong tendency in Swahili, and indeed in most Bantu languages (Kanerva, 1990; Hyman & Katamba, 1990). This stress rule is exemplified below, where primary stress in (7a) is on the verb *pig-*, the penultimate syllable. In (7b), with the addition of the applicative suffix, stress moves rightward onto the applicative suffix, which is now the penultimate syllable. In (8a), stress is on the penultimate syllable of the word *maktába*, a loan word from the Arabic, meaning ‘library’, and with the addition of the locative suffix, stress moves rightward onto the penultimate syllable.

- (7a) morphological: ni - na - m - píg - a
 SA_{1s}-pres - OA_{3s}-hit - IND
 syllabic: ni - na - m - pí - ga
 ‘I am hitting him’
- (7b) morphological: ni - na - m - pig - í - a
 SA_{1s}-pres- OA_{3s}- hit - appl- IND
 syllabic: ni - na - m - pi - gí - a
 ‘I hit him (for someone/with something)’
- (8a) maktába
 library
- (8b) maktabá-ni
 library -locsuff
 ‘In/to the library’

³ However, speakers of Nairobi Swahili tend to shift between the standard *lázima* and the more colloquial *lazíma*. Vitale (1985) shows how this shifting between non-penultimate stress and penultimate stress for loan words is common even in standard dialects, suggesting that the shift from non-penultimate stress to penultimate stress is one that takes time. Such words as *lazíma* are new loan words which are in the midst of being incorporated into the phonological system of Swahili, supporting the prominence of the penultimate stress rule in Swahili.

Additionally, secondary stress usually occurs in the verbal complex on the SA marker, as in (9a) below. Barrett-Keach (1986) argues that there is phonological word boundary after the T marker, and that secondary stress on subject agreement is simply a product of the penultimate stress rule in Swahili. She was concerned with arguing for a separate AUX node in the syntax of Swahili.

She provides evidence that if the T marker is made longer, then secondary stress moves rightward so as to fall on the penultimate syllable. In (9b), the tense marker is disyllabic, and secondary stress falls on the first syllable of the tense marker. In (9c), ‘mekwisha’ is a trisyllabic marker with secondary stress falling on the medial syllable. In all cases, secondary stress falls on the penultimate syllable from the right edge of the T marker. She concludes that this rightward movement of secondary stress is because there is a prosodic word boundary between T and the rest of the verbal complex, a fact that is consistent with AUX forming a constituent.

- (9) a. Nì - me - fík - a
 SA_{1s}-pr.perf.-arrive-IND
 ‘I have arrived.’
- b. Ni - mèsha - fík - a
 SA_{1s}-perf.comp.-arrive-IND
 ‘I have already arrived.’
- c. Ni - mekwìsha - fík - a
 SA_{1s}-perf.finish-arrive-IND
 ‘I have finished arriving.’

Thus primary stress falls on the penultimate syllable of the verbal complex, and secondary stress falls on the penultimate syllable from the right edge of the T marker.

Swahili has a bias towards trochaic feet (Strong-Weak) as opposed to iambic feet (Weak-Strong). This is evident from the penultimate stress rule, as well as loan-word adaptations. Penultimate stress has been argued to bias languages such as Italian and Spanish towards trochaic feet (see Hayes, 1991). Furthermore, loan words that have iambic feet are adapted in ways that differ from loan words that have trochaic feet. For example, the loan word *kitaab* [kita:b] (W-S) from the Arabic ‘book’ has been adapted to [kitábu] to incorporate a trochaic foot. Thus the original iambic foot of the Arabic [kita:b] is modified by the addition of a final vowel, making the structure into one that contains a trochaic foot (W [S-W]). Similarly, the Arabic word *hilaal* [hila:l] (W-S) meaning ‘crescent’ has been adapted to [hiláli], again incorporating a trochaic foot. The process involves the insertion of a final vowel, with stress remaining on the original stress-bearing segment.

This could be argued to simply be insertion of a final vowel for the sake of making the stress bearing segment the penultimate syllable, and not a preference for trochaic feet per se. However, in other loan words that include a trochaic foot, the stress can be seen to optionally move to the penult, e.g., the Arabic *lázim* is adapted into the Swahili *lázima*. A final vowel is added to the adapted version because Swahili does not allow coda consonants, but this has no bearing on the issue of stress. This pronunciation of the word (with antepenultimate stress) is in free variation with the penultimate stress form: *lazíma* (see Vitale, 1992; 1985). Thus in the antepenultimate stress form, stress remains in a marked position (antepenultimate) but within a trochaic foot. The word is slowly being adapted further by stress moving rightward into the penultimate position, hence the option to pronounce the word with penultimate stress as *lazíma*.

This shows that non-penultimate stress is tolerated if it occurs within a trochaic foot, but not if it occurs within an iambic foot. Another such example is *héshima*, from the Arabic [hëshma] meaning ‘honor, dignity’. The standard pronunciation of this word is *hëshima*, with antepenultimate stress. However, the penultimate stress form also occurs: *heshíma*. Thus stress on the antepenultimate syllable is tolerated within a trochaic foot, but not within an iambic foot. I interpret these facts as suggesting that Swahili has a bias towards trochaic feet. We will return to these facts in section 2.6 in our discussion of whether T is an auxiliary verb or a T marker. It will also be relevant in chapter 4 when we discuss the Metrical Omission Model (Gerken 1991).

No statistical corpus analysis has been reported that can verify this claim that there is a trochaic bias in Swahili, but the penultimate stress rule along with the facts on loan-word adaptation are indicative of this conclusion. Furthermore, Demuth (1994) claims that Sesotho is a language that has a bias towards trochaic feet (p.128), and she uses this fact to apply a metrical model of omission to child omission of noun class prefixes. We will also apply these facts for Swahili to child Swahili in chapter 4, where we discuss the Metrical Omission Model (Gerken 1991).

2.3 Noun Classes

As is typical of most Bantu languages, Swahili has a large number of noun classes. A noun class can be thought of as being similar to the gender systems found in the Romance and Germanic languages, in that it is an arbitrary lexical feature. The noun class system treats singular and plural nouns as distinct noun classes (Meinhoff, 1932), as can be seen in table 2.1 below (see Carstens, 1991; Bresnan & Mchombo, 1989 for a discussion).

For example, noun class 1 and noun class 2 are considered distinct noun classes, they refer to the same object, with noun class 2 ('people') being the plural counterpart of noun class 1 ('person').

In all examples I will gloss nouns by the numbers in table 2.1.

Thus the noun in example (10) is of noun class 8.

- (10) vi – kombe
8–cup
'cups.'

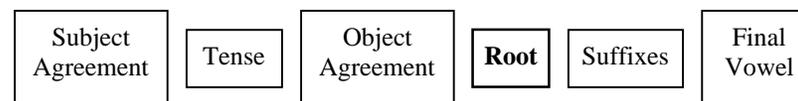
Table 2.1 Meinhoff's Noun Classification System

| Class | Example | Gloss |
|-------|-----------|----------------|
| 1 | m-tu | person |
| 2 | wa-tu | people |
| 3 | m-ti | tree |
| 4 | mi-ti | trees |
| 5 | gari | car |
| 6 | ma-gari | cars |
| 7 | ki-tabu | book |
| 8 | vi-tabu | books |
| 9 | n-yumba | house |
| 10 | n-yumba | houses |
| 11 | u-bao | board |
| 14 | u-kweli | truth |
| 15 | ku-kimbia | to run |
| 16 | mahali | specific place |
| 17 | mahali | general place |
| 18 | mahali | inside place |

2.4 The Verbal Complex

Recall that the order of the elements in the verbal complex is as shown below in (11):

- (11) Order of elements in the verbal complex:



Of these elements, subject agreement, tense and the final vowel are the only ones which are obligatorily present with the root in every affirmative Swahili utterance⁴. I will describe the structure, occurrence and function of each of these elements in turn (beginning with Subject agreement and ending with the final vowel).

2.4.1 Subject agreement paradigm

The subject agreement marker is obligatory in almost all contexts in Swahili, whether the subject is overtly present or not. Examples (12)-(13) exemplify this:

- (12) m-toto m-zuri a – me – anguk - a
1-child 1-good SA_{3s} – Pr.perf. – fall - IND
'The good child has fallen.'
- (13) wa-toto wa-zuri wa – me – anguk - a
2-child 2-good SA_{3pl} – Pr.perf. – fall - IND
'The good children have fallen.'

In example (12), the subject is *mtoto* 'child'. The subject agreement marker on the verb ('a-') agrees with the subject in noun class. It is glossed as SA_{3s} (as opposed to SA₁) because all nouns in classes 1 and 2 agree in

⁴ Although see section 2.4.1 for a description of a non-standard colloquial dropping of SA.

person/number. In example (13), the subject is of noun class #2, and the subject agreement marker changes appropriately⁵.

When the subject of the sentence is a personal pronoun (as opposed to a lexical item that carries class features), the paradigm of agreement exhibited on the verb is not homophonous with any noun class. Rather, agreement occurs with the subject in person and number.

(14)

| | <u>Optional Pronoun</u> | <u>SA</u> | <u>T</u> | <u>V</u> | <u>FV</u> | <u>Gloss</u> |
|--------------------------|-------------------------|-----------|----------|----------|-----------|----------------|
| 1 st singular | Mimi | ni- | -li- | -anguk- | -a | I fell |
| 2 nd singular | Wewe | u- | -li- | -anguk- | -a | You fell |
| 3 rd singular | Yeye | a- | -li- | -anguk- | -a | He fell |
| 1 st plural | Sisi | tu- | -li- | -anguk- | -a | We fell |
| 2 nd plural | Ninyi | mu- | -li- | -anguk- | -a | You (all) fell |
| 3 rd plural | Hawa ⁶ | wa- | -li- | -anguk- | -a | They fell |

In (14) above, the personal pronoun is generally omitted in such a sentence, but can occur for emphasis. Note that there is a separate agreement morpheme corresponding to each person, e.g., *ni-* corresponding to first person, singular; *u-* corresponding to second person, singular, etc.

2.4.1.1 Nairobi Swahili Vs. Standard Swahili

Returning to the difference in agreement between Standard Swahili and Nairobi Swahili, I mentioned in the previous section that Nairobi has a

⁵ In addition to subject verb agreement, these examples show agreement within the subject adjectival phrase. The agreement within the adjectival phrase is agreement between the head noun and the modifying adjective (see Carstens (1991) for a detailed analysis of Swahili DPs and agreement within DPs).

⁶ This is specifically Nairobi Swahili. The 3rd person plural pronoun in Standard Swahili is *wao*.

reduced agreement pattern. I will first describe this phenomenon, and then I will provide examples from the adults in the corpus of data. Speakers of Nairobi Swahili use a full set of agreement markers when referring to a subject of noun classes 1 and 2. Recall that these noun classes are animate (singular and plural, respectively). The agreement markers are the ones listed in (14), with lexical nouns taking either *a-* or *wa-* as the agreement marker, and pronouns taking the full paradigm in (14). However, for all other classes, *i-* and *zi-* are the agreement markers, with *i-* being singular and *zi-* being plural. In Standard Swahili there are different agreement markers for each of the first 10 classes (see table 2.2). Examples (15) – (17) below show sample utterances from Nairobi speakers, taken from the audio recordings in this corpus. In (15), we see an example of a singular subject from noun class 7. In standard Swahili, the agreement marker on the verb is of the form ‘*ki-*’. However, notice that in Nairobi Swahili the agreement marker is *i-*. Similarly in (16), the subject is from noun class 8, and Nairobi Swahili uses *zi-* rather than the standard *vi-*. In (17), Nairobi Swahili uses the same plural *zi-*, rather than the standard *ya-*. Table 2.2 lists the Standard Swahili subject agreement markers with the corresponding Nairobi Swahili subject agreement markers.

| | <u>Standard Swahili</u> | <u>Nairobi Swahili</u> |
|------|---|--|
| (15) | Ki-tanda ki – me – vunj – ik – a 7-bed SA ₇ -pr.prf.- break- state-IND ‘The bed has broken’ | Ki-tanda i – me – funj – ik – a 7-bed SA _{pl} -pr.prf.-break-state- IND ‘The bed has broken’ |
| (16) | Vi-tabu vi – me – anguk – a 8-book SA ₈ -pr.prf..-fall-IND ‘The books have fallen’ | Vi-tabu zi – me – anguk – a 8-book SA _{pl} -pr.prf.-fall-IND ‘The books have fallen’ |
| (17) | Ma-gari ya – li – ingi – a 6-car SA ₆ -past-enter-IND ‘The cars entered’ | Ma-gari zi – li – ingi – a 6-car SA _{pl} -past-enter-IND ‘The cars entered’ |

Table 2.2 Standard Swahili and Nairobi Swahili subject agreement morphology

| Noun Class | Standard Swahili | Nairobi Swahili |
|------------|------------------|-----------------|
| 1 | a | a |
| 2 | wa | wa |
| 3 | u | i |
| 4 | i | zi |
| 5 | li | i |
| 6 | ya | zi |
| 7 | ki | i |
| 8 | vi | zi |
| 9 | i | i |
| 10 | zi | zi |
| 11 | u | ? |
| 15 | ku | ? |

Nairobi Swahili speakers have a reduced agreement system, marking animacy (noun classes 1 and 2) and number, but not noun class itself. However, since many adult speakers of Nairobi Swahili are also schooled in Standard Swahili, and because of the stigma associated with Nairobi Swahili, getting judgments about this phenomenon is extremely difficult. The only evidence that exists comes from spontaneous speech such as that found in the adult data collected for this study. In this corpus there were no occurrences of nouns of classes 11 or 15, and so it is unclear what the Nairobi Swahili subject agreement markers would be for these classes.

2.4.1.2 SA Omission

According to traditional grammars, subject agreement (henceforth SA) is obligatory in all contexts, e.g., Ashton (1947): ‘In Swahili the verb cannot stand alone as in English, but must be prefixed by the Pronominal Concord proper to the noun which forms its subject’ (p.15)⁷. Similarly,

⁷ Pronominal Concord = subject agreement

Myachina (1981) claims, ‘The subject markers...are an indispensable component of the verbal complex’ (p.49). Vitale (1981) is just as unequivocal, ‘The important fact is that the subject affix is an obligatory morphological category of the verb’ (p.15). Krifka (1995), in his survey of Swahili syntax, is a little more cautious, but nevertheless claims that it is obligatory in all cases except in certain tenses, the habitual *-hu-* and the allomorph of the present tense *-a-* (p.1399). A survey of adult speech in the corpus used in this study reveals that this is not entirely true in this dialect of Swahili. Adult speakers of Nairobi Swahili overwhelmingly use SA in verbal contexts, but occasionally omit SA. This omission does not correspond to the use of habitual *-hu-* or present tense *-a-*, but rather is used in all tense contexts. The discourse context for this omission appears to be in cases when the subject is extremely salient, and when the topic and the subject are co-referent. Of all verbal utterances in the adults’ speech, 5% of their utterances were missing SA. While this is a small proportion of omitted SA markers, as I will discuss later in the syntax section of this chapter, omission of SA is considerably more frequent than omission of any other obligatory affix, for example, tense or mood, omission of which is nonexistent. So the omission of SA constitutes a genuine grammatical possibility for adults⁸. An analysis of this sentence type will be provided later in this chapter.

⁸ In fact Ashton (1947) notes that SA may be omitted in certain clauses that take the *ka* tense marker. She claims that the omission of SA results in an emphatic or surprised connotation. Scotton (1969) also describes cases in which up-country Swahili speakers and Bagandan Swahili speakers frequently omit SA. See below for details.

2.4.2 Tense/Aspect

In adult Swahili, every indicative utterance contains a tense/aspect marker. Unlike SA, it is ungrammatical to omit the tense/aspect marker in any utterance even if the temporal frame/reference is clear from discourse or other sources. Table 2.3 below shows the various tense/aspect markers that occur in Swahili:

Table 2.3 Tense/aspect morphemes in Swahili

| Tense/Aspect Morpheme | Meaning |
|-----------------------|----------------------------|
| li | past |
| na | present on-going/habitual |
| ta | future |
| ka | Narrative, resultative |
| me | present perfect |
| sha | present perfect completive |
| ki | conditional |
| nga | hypothetical |
| ku | infinitival |

li is used in simple past tense contexts. It is an absolute tense (Comrie, 1976), that is, it can be used as an anchoring tense in discourse and is not dependent on the surrounding context. The future tense marker is a relative tense marker, in that it takes its reference from the immediately preceding context or the matrix tense when in an embedded clause. Similarly, *ka*, *me* and *sha* are all relative tense markers. *Ka* is referred to by Ashton (1947) as a consecutive marker. I will refer to it as a continuative marker, meaning that it takes an anaphoric interpretation from the previous action. A verb marked with the *ka* marker describes an action that occurred immediately after the previously mentioned action. This marker is most frequently used in narratives and story telling in order to

drive the story line forward in time. The second clause in example (18) is marked with *ka*.

- (18) a – li – ruk – a chini a – ka – kimbi – a
 SA_{3s}–past–descend–IND down SA_{3s}–cont–run–IND
 ‘He climbed down, (and then) he ran off.’

Me and *sha* are very similar in meaning, with the only difference being that *sha* carries a sense of completion. They are both the present perfects of result. In English translations of Swahili text, *sha* is often translated as "have already done X", whereas *me* is often translated as "have done X".

- (19) a. ni – me – kul – a
 SA_{1s}–perf–eat–IND
 ‘I have eaten.’
 b. ni – sha – kul – a
 SA_{1s}–perf.comp.–eat–IND
 ‘I have already eaten.’

The next two tense morphemes are somewhat less common in Nairobi Swahili. *Nga* is a hypothetical morpheme, and *ki* is a habitual or conditional marker. *Nga* is considerably less common than any of the other tense affixes, and was completely absent from any adult or child utterances in the corpora used in this study.

Finally, there is the infinitival *ku* marker. This occurs in two contexts: first, as the complement of a control type verb, and second, as a gerund. This second function of the infinitive was seen earlier in the description of noun classes, where noun class 15 was the infinitival noun class. (20) and (21) are examples of each of these two functions of the infinitive marker:

(20) Ni - na - tak - a ku - ondok - a sasa → True Infinitive
 SA_{1s}-pres-want-IND INF-leave-IND now
 'I want to leave now.'

(21) Ku - imb-a ku - me - kwish - a → Gerundive Infinitive
 INF -sing -IND SA₁₅-pr.prf.-end - IND
 'The singing has ended.'

The fact that Swahili has an infinitival form will be of significance when we investigate the question of root infinitives (RIs) in the speech of Swahili children in chapter 4, since the existence of this morpheme means that the language does have the potential to exhibit RIs.

2.4.3 Negation

Following is a simplified description of negation in Swahili (see Ashton, 1947; Vitali, 1981; Krifka, 1995 for a full description). Negation occurs in three positions in the sentence: 1) at the head of the verbal complex, 2) within the verbal complex, and 3) at the end of the verbal complex in the form of a negative final vowel. These three reflexes of negation are not mutually exclusive (see below), as most negative sentences require at least two of these negation positions to be overtly expressed. In the simple future tense, the negative prefix *ha* attaches to the complex before the SA marker, as in (22b) below.

(22) a. a - ta - nunu - a vi - tabu
 SA_{3s}-fut.- buy - IND 8-book
 'She will buy (some) books.'

(22) b. H - a - ta - nunu - a vi - tabu
 neg- SA_{3s} - fut - buy - IND 8-book
 'He will not buy (some) books'

In this case, the only overt representation of negation is the negative marker at the beginning of the verbal complex. However, in other tenses the two other negative positions are activated. Table 2.4 below presents the negation paradigm⁹:

Table 2.4 Swahili Negation Paradigm

| Tense (Morpheme) | Negative Paradigm | Example |
|--------------------------------|-------------------|--|
| past (li) | ha-SA-ku-V-IND | Ha - tu - ku - kul- a ma - embe neg-SA _{1pl} -past-eat-IND 6-mango 'We did not eat mangoes.' |
| Present (na) | ha-SA-V-NEG | Ha - tu - kul - i ma - embe neg - SA _{1pl} - eat - NEG 6-mango 'We don't eat mangoes.' |
| future (ta) | ha-SA-ta-V-IND | Ha - tu - ta - kul - a ma - embe neg - SA _{1pl} - fut - eat - IND 6-mango 'We will not eat mangoes.' |
| present perfect (me) | ha-SA-ja-V-IND | Ha - tu - ja - kul - a ma - embe neg - SA _{1pl} - pr.prf-eat-IND 6-mango 'We have not eaten mangoes.' |
| present perfect complete (sha) | ha-SA-ja-V-IND | Ha - tu - ja - kul - a ma - embe neg - SA _{1pl} - pr.prf-eat-IND 6-mango 'We have not eaten mangoes.' |

In the case of the past *li*, the negative form is a negative correlative, i.e., *hu*, appears before the SA marker and the suppletive form *ku* of the tense marker *li* occurs in tense position.¹⁰ For the present tense, when *ha* appears before SA, we find the verb marked with a negative final

⁹ The tenses which are not shown in this table (narrative, hypothetical, habitual/conditional, and infinitive) form their negation through a complex negation. A cleft-construction is used which clefts an auxiliary *be* verb plus negation, and this forms the only mechanism for negation in these circumstances. An appropriate translation of such a construction would be "It was not that..."

¹⁰ Note that this negation particle is homophonous with both the 2nd person singular Object Agreement Marker as well as the Infinitive marker.

vowel. As mentioned above, for the future, *ha* appears before SA, with the regular future tense marker remaining the same. For the present perfect and the present perfect completive, similar to the past perfective negative, the *ha* marker appears before SA, and suppletion of the tense marker results in *ja* appearing in the regular tense position. In summary, in addition to the left-most negative marker (*ha*), there are two other positions in which negation surfaces: suppletion of the tense marker in the past, present perfect, and present perfect completive tenses, and the negative final vowel in the present tense.

The children in this study used syntactic negation of the sort described in this section relatively infrequently. Negative utterances were formed with lexical negation which is non-adult like, or simply *hapana* ‘no’. However, there were not enough tokens of negation for a comprehensive investigation into this phenomenon, and must be left for future investigation.

2.4.4 Object Agreement Marking

Object agreement in Nairobi Swahili is similar to subject agreement, in that it has been reduced from the Standard Swahili paradigm to mark animacy and number only. Table 2.5 below shows the full paradigms of Standard Swahili and Nairobi Swahili. Examples follow the tables. Personal pronoun OA is no different in Nairobi Swahili from Kiswahili Sanifu, and this is presented in table 2.6.

The Object agreement marker is obligatory in simple transitive sentences when the object is specific and/or animate¹¹. When the object is non-specific and inanimate, object agreement is obligatorily absent. We will return to this point in section 2.7.

Table 2.5 Standard and Nairobi Swahili Object Agreement Paradigm for noun classes

| Noun Class | Standard Swahili | Nairobi Swahili |
|------------|------------------|-----------------|
| 1 | m | m |
| 2 | wa | wa |
| 3 | u | i |
| 4 | i | zi |
| 5 | li | i |
| 6 | ya | zi |
| 7 | ki | i |
| 8 | vi | zi |
| 9 | i | i |
| 10 | zi | zi |
| 11 | u | i/u |
| 15 | ku | ku/i |

Table 2.6 Personal pronoun Object Agreement

| Person | Object Agreement |
|--------------------------|------------------|
| 1 st singular | ni |
| 2 nd singular | ku |
| 3 rd singular | m |
| 1 st plural | tu |
| 2 nd plural | mu ¹² |
| 3 rd plural | wa |

¹¹ See Bresnan & Mchombo (1987) for a discussion of the object agreement marker in Bantu, and Moshi (1985) and Ngonyani (1996) for a specific discussion of object agreement in Kiswahili.

¹² In Standard Swahili, the OA for 2nd plural is *wa*, not *mu*.

- (23) ni – ta – **zi** – nunu– a nguo
SA_{1s}–fut–OA₁₀–buy–IND clothes
'I will buy the clothes.'
- (24) A – li – **ni** – pig - a
SA_{3s}–past–OA_{1s}–hit–IND
'He hit me.'
- (25) A – li – **ku** – pig - a
SA_{3s}–past–OA_{2s}–hit–IND
'He hit you.'
- (26) A – li – **m** – pig - a
SA_{3s}–past–OA_{3s}–hit–IND
'He hit him'¹³

2.4.5 Roots

The verbal root in Swahili, and indeed in most Bantu languages, is generally monosyllabic. Not all verb roots are monosyllabic however, and the majority of the multi-syllabic verb roots are verbs which were originally borrowed from other languages (most commonly from Arabic). The verb root is generally CVC, although other structures are not uncommon¹⁴:

¹³ The reference of the object him cannot be the subject him, i.e., this cannot take an anaphoric interpretation. For an anaphoric interpretation the anaphoric marker *-ji-* must be used in preverbal position instead of the 3rd person singular marker *-m-*.

¹⁴ In this table, the roots are given along with the final vowel, which in Swahili is obligatory.

Table 2.7 Some root structures for Swahili verb roots

| Root structure | Example + Final Vowel | Gloss |
|----------------|-----------------------|---------------------|
| C | p-a | give |
| VC | iv-a | become ripe, mature |
| CVC | pig-a | hit |
| CVCC | shind-a | win |
| VCVC | azim-a | borrow, lend |
| VCC | amb-a | adhere |
| VCCVC | anguk-a | fall |

The smallest verb root attested in Swahili consists of a single consonant or a consonant cluster. There are very few such verbs, although they are some of the most commonly used verbs in the language:

Table 2.8 monosyllabic verb roots and their citation forms

| Verb Root | Root+IND | Citation Form | Meaning |
|-----------|----------|---------------|---------|
| -l- | l-a | kula | eat |
| -ny- | ny-a | kunya | drop |
| -nyw- | nyw-a | kunywa | drink |
| -f- | f-a | kufa | die |
| -j- | j-a | kuja | come |

Each verb root obligatorily takes a mood final vowel (shown in the second column), forming the minimal verb stem for these monosyllabic verbs. Because stress falls on the penultimate syllable in Swahili, a syllable is inserted before these verb roots to make them disyllabic. Thus the citation form for these verbs is shown in the third column. Notice that this syllable is homophonous with the infinitive marker. I distinguish between these two by referring to the true infinitive as 'infinitive *ku*' and the syllable inserted for phonological purposes as 'dummy *ku*'.

In Standard Swahili this dummy *ku* occurs only in certain contexts where the prefix cannot carry stress, e.g., in a tensed clause (27). Certain

prefixes may carry stress, in which case the *ku* is not inserted (28a).

However, these monosyllabic verbs have been reanalyzed in Nairobi

Swahili as disyllabic verb stems in which *ku* is part of the verb root (28b):

(27) a – na – ku – l – a Standard and Nairobi Swahili
 SA_{3s}–pres–ku–eat–IND
 ‘He is eating.’

(28) a. Juma hu – l – a Standard Swahili
 Juma hab–eat–IND
 ‘Juma (habitually) eats.’
 b. Juma hu – kul – a Nairobi Swahili
 Juma hab.–eat–IND
 ‘Juma (habitually) eats.’

This phenomenon will become relevant when we discuss the syntax of the tense prefix in section 2.6

2.4.6 Suffixes

As mentioned earlier, in addition to the verbal prefixes, there are a host of suffixes. Because the focus of this study is the acquisition of prefixes, this section is provided simply as additional information for the reader. There can be up to three suffixes attached to the verb, in addition to the final vowel (see next section). This will not be discussed in great detail; for a more in-depth discussion of the ordering of suffixes, the reader is directed to Krifka (1995) and Vitale (1981). In table 2.9 below is a list of the five most common Swahili verbal suffixes, as discussed in Ashton (1947). See Ashton (1947), Myachina (1981) for discussion of other suffixes.

Table 2.9 Common Verbal Suffixes in Swahili

| Suffix Function | Description |
|-----------------|---|
| Passive | Causes the theme to be the subject of the sentence |
| Applicative | Introduces an additional argument to the sentence (see below) |
| Stative | Makes the verb stative |
| Causative | Makes the verb causative |
| Reciprocal | Reciprocalizes the verb |

The passive suffix is extremely common in Swahili. The passive in Swahili works in a manner similar to other languages in that the subject of the sentence is the theme/patient of the action, with passive morphology occurring on the verb. The passive suffix is [w], as in the following example:

(29) a. A – li – ni – pig – a ACTIVE
 SA_{3s} - past perf.-OA_{1s} -hit - IND
 ‘He hit me.’
 b. Ni - li -pig - w - a (na yeye) PASSIVE
 SA_{1s}–past–hit–passive–IND (by him)
 ‘I was hit (by him).’

The object is raised to subject position, and the logical subject may appear optionally in a ‘na’ phrase at the end of the sentence.

The applicative suffix introduces an additional argument into the sentence. Ngonyani (1996) shows that in Swahili the applicatives can be categorized into three broad groups, with further sub-groupings possible. Below in table 2.10 is a full list of Ngonyani’s groupings:

Table 2.10 Types of Applicative sentences in Swahili

| Benefactive-Type | Instrumental-Type | Locative-Type |
|------------------|-------------------|---------------|
| Benefactive | Instrumental | Locative |
| Goal | Motive | |
| Malefactive | | |

The example (30) shows a non-applicative sentence, and (31) shows its applicative counterpart. (31) is a benefactive applicative, while (32) is an instrumental applicative and (33) is a locative applicative¹⁵.

- (30) Juma a - li -nunu-a vi-tabu
 Juma SA_{3s}-past-buy-IND 8-book
 'Juma bought books.'
- (31) Juma a - li - wa - nunu - li - a wa-toto vi-tabu BENEFACTIVE
 Juma SA_{3s}-past-OA_{3pl}-buy-APPL-IND 2-child 8-book
 'Juma bought the children books.'
- (32) Mawe, wa - li - vunj - i - a ch-ungu INSTRUMENTAL
 6-stone, SA_{3pl}-past-break-APPL-IND 7-pot
 'The stones, they broke the pot with them.'
- (33) Wa-teja wa - li - l - i - a ch-akula ofisi - ni LOCATIVE
 2-customer SA_{2pl}-past-eat-APPL-IND 7-food office-LOC
 'The customers are food in the office.'

Sentence (30) is a non-applied sentence, with two arguments: the subject *Juma* (a proper name) and the direct object *vitabu* 'books'. In sentence (31), the applicative suffix *-li-* is attached to the verb, and there is an additional argument added: the benefactive object *watoto* 'children'. Notice that object agreement is obligatory, with agreement being with the applicative object. For an excellent and detailed description of the

¹⁵ The locative example is Ngonyani's example, and so the verb root is given as [l]. In Nairobi Swahili the verb has incorporated the dummy infinitive into the verb root, resulting in [kul].

differences between the various types of applicatives in Swahili, see Ngonyani (1996).

In Swahili there is a suffix for stative verbs. A verb can take the stative suffix *-ik-* and its interpretation is what traditional grammarians describe as the Middle Voice.

- (34) Ni - me - vunj - a ki-kombe
 SA_{1s}-pr.prf.. - break - IND 7-cup
 'I have broken a cup.'
- (35) Ki-kombe ki - me -vunj - ik - a STATIVE
 7-cup SA₇- pr.prf.-break- STATE - IND
 'The cup has/is broken.'

Causatives are marked with the suffix *-ish-* as in examples (36-37) below. Example (36) is a non-causative, transitive verb. Example (37) is the same verb with a causative suffix.

- (36) Wa-toto wa - na - imb - a.
 2-child SA_{3pl}- pres- sing - IND
 'The children are singing.'
- (37) M-walimu a - na - wa - imb - ish - a wa-toto CAUSATIVE
 1-teacher SA_{3s}-pres-OA_{3pl} -sing-causative-IND 2-child
 'The teacher is making the children sing.'

The final suffix to be exemplified is the reciprocal suffix, as in

- (39).
 (38) Mariam a - li - m - pig - a Juma
 Mariam SA_{3s}- past perf.- OA_{3s} - hit - IND Juma
 'Mariam hit Juma.'
- (39) Mariam na Juma wa - li - pig - an - a RECIPROCAL
 Mariam and Juma SA_{3pl}- past-hit-RECIP - IND
 'Mariam and Juma hit each other.'

The remaining suffixes are the conversive suffix (reverses the meaning of a verb, e.g., fold → unfold), augmentative suffix (intensifies the meaning of a verb), the inceptive suffix (indicates a state entered upon), etc.

For a full description of the suffixes in Swahili, refer to Ashton (1947); Krifka (1995); Vitale (1981), Myachina (1981).

2.4.7 Final Vowel

Unlike the other suffixes described in the previous section, the final vowel in Swahili is obligatory. In Swahili and other Bantu languages, the final vowel has generally been described as a mood vowel. Bresnan & Mchombo (1987) and Bresnan & Kanerva (1989) gloss the final vowel *a* as "IND" (indicative) as I have done here. However, there is some variation in the function of the final vowel in some other Bantu languages, where it seems to have no interpretable meaning (e.g., Mutaka (1994) claims the final vowel in Kinande (an Eastern Congolese Bantu language) "has no meaning", p.33). Kinyalolo (1991) disagrees with the analysis that the final vowel is a mood marker. He suggests that while it is clear that *e* is a subjunctive marker, in KiLega (another Eastern Congolese Bantu language) it also has other meanings. For example, it may also mark immediate future (as opposed to the predictive future, which is more compatible with a subjunctive interpretation). Kinyalolo proposes that *a* is a default vowel that occurs whenever there is no other tense/aspect/mood morpheme to carry the appropriate features (p.304).

In Swahili and most other Bantu languages, there are three forms of the final vowel: [a], [e] and [i]. I will continue to make a distinction between indicative and subjunctive final vowels, since in Swahili, unlike KiNande or KiLega, this morphological distinction is associated with the corresponding semantic distinctions. I will gloss the *a* final vowel as IND and the *e* final vowel as SUBJ. The negative final vowel has already been discussed, and will be glossed as NEG.

The indicative in Swahili is generally associated with realis-type meaning. A realis context can be defined as one in which the event/state has been experienced, while an irrealis context is one in which the event/state has not been experienced (Bybee *et al.*, 1994; Givón, 1995; Palmer, 1987). Table 2.11 below shows the various meanings and contexts in which the indicative and subjunctive final vowel are used.

Table 2.11 Contexts for use of final vowel

| Indicative | Subjunctive |
|------------------|---------------------|
| Past | Request |
| Present perfect | Express desire/wish |
| Present simple | Preference |
| Present habitual | Obligation |
| Future | Probability |
| Infinitive | Disbelief |
| Imperative | |

IND and SUBJ in Swahili map straightforwardly onto realis and irrealis, with the exception that the future is marked as indicative. SUBJ carries meanings associated with epistemic and deontic modality (Palmer, 1979, 1986; Givón, 1995), while IND is associated with realis assertions (with the exception of the future). The fact that the future is the one context which crosses the divide between irrealis and realis is not surprising, according to Givón (1995). Givón argues that there is a continuum of meanings, ranging from prototypically realis to prototypically irrealis. He claims that on the realis side are meanings such as past, present etc., which cross-linguistically strongly tend to be encoded with indicative morphology. On the other end of the continuum are the irrealis meanings such as desire, obligation, necessity, permission, etc. which strongly tend to be encoded with subjunctive morphology. In between these two extremes are middle

categories which are subject to cross-linguistic variation. One such ‘middle’ category is the future, and thus the fact that Swahili marks the future with indicative morphology is unsurprising. It is important to keep in mind that this is an irregularity in the linguistic system which learners must acquire. For a discussion of the acquisition of mood, see Deen & Hyams (2002).

As table 2.11 above shows, all tensed clauses occur in the indicative. Furthermore, subjunctive requires that tense be absent. Subjunctive occurs either in embedded clauses, or on main verbs with the meanings noted in table 2.11. Following are examples of sentences with subjunctive final vowels.

- (40) tafadhali ni - pat - i - e ch - ai REQUEST
 please OA_{1s}-give-APPL-SUBJ 7-tea
 ‘Please give me (some) tea’
- (41) Ni - na - tak - a a - nunu - e nguo mpya DESIRE/WISH
 SA_{1s}-pres-want-IND SA_{3s}-buy-SUBJ clothes new
 ‘I want (that) he buy new clothes’
- (42) Lazima u - ni - imb - i - e wimbo OBLIGATION
 must SA_{2s}-OA_{1s}-sing-APPL-SUBJ song
 ‘You (really) must sing me a song’

2.4.8 Imperatives

Imperatives in Swahili prototypically do not take an overt subject unless (as in other languages) the subject is focused or contrastive. There is no imperative mood marker, as the imperative takes the indicative final vowel. The imperative does not occur with SA (43b) or T (43c). A full range of suffixes is also possible with the imperative (44 a-b).

- (43) a. nunu - a m - kate
 buy - IND 3 - bread
 ‘Buy bread!’
- b. * u - nunu - a m - kate SA is ungrammatical
 SA_{2s} - buy - IND 3 - bread in imperatives
- c. * ta - nunu - a m - kate Tense is ungrammatical
 fut - buy - IND 3 - bread in imperatives
- (44) a. imb - ish - a wa - toto Causative in imperative
 sing - caus. - IND 2 - child
 ‘Make the children sing!’
- b. nunu - li - a wa - toto m - kate Applicative in imperative
 buy - APPL - IND 2 - child 3 - bread
 ‘Buy the children some bread!’

Often, the difference between subjunctive and indicative can be subtle and depends on speaker intention. For example, the sentence in (40) is a request, and has a subjunctive final vowel. If the speaker intended to be more forceful, or to signal social dominance over the addressee, the indicative would be used, turning the sentence into an imperative:

- (44) ni - pat - i - a ch - ai IMPERATIVE
 OA_{1s} - give - APPL - IND 7 - tea
 ‘Give me (some) tea’

In example (44) there is an object agreement marker, which in Nairobi Swahili is required by the specific applicative object (see section 2.4.6 on the applicative)¹⁶. In cases where there is no applicative, no object agreement is required, and so the imperative form surfaces as a bare stem.

¹⁶ OA is usually absent in imperatives in Standard Swahili.

Important for our purposes is that the imperative is a bare stem (i.e., the verb root and a mood final vowel) – the only case in adult Swahili in which SA and T are ungrammatical and the bare stem is fully grammatical. This will be of consequence when we consider acquisition. As we will see in chapter 4, because the frequency of imperatives in child-directed speech is high, the children are hearing bare forms relatively frequently. Thus the use of bare forms by children is not unexpected.

- **Syntax of Swahili**

In the first part of this chapter we discussed the morphological system of Swahili, paying particular attention to the verbal morphology. In this section I turn to the syntax of Swahili, focusing on issues that will be of relevance in later chapters. In general I use terminology that is pre-Minimalist, although most of what I claim can be rendered in more modern terms. The organization of this section is as follows. First, I discuss a topic in Bantu linguistics that has recently received much attention: the identity of SA (section 2.5) and the identity of tense (2.6). Specifically, I discuss whether SA in Nairobi Swahili is actually agreement (as I have described it in section 2.4.1), or whether it is better analyzed as a pronoun. I also discuss whether tense in Swahili is actually tense marking (as an inflectional prefix), or whether it is better analyzed as an auxiliary verb. The discussion is extended to OA in section 2.7. I conclude that SA and OA in Swahili are in fact agreement morphemes, as described earlier, but that the status of tense is somewhat less clear. The discussion demonstrates how theoretical studies cannot always conclusively answer such questions,

and in subsequent chapters we will find that the acquisition data make significant contributions to these theoretical debates.

In section 2.8, I derive a simple tensed clause. In the final section of this chapter, 2.9, I introduce a construction in Swahili that has not been recognized in the literature thus far, [-SA] clauses (clauses in which SA has been omitted). I show that adults omit SA in restricted contexts, and that such clauses have syntactic characteristics that differentiate them from full clauses. These two types of clauses provide evidence of two different types of null subject, each of which the child must recognize and acquire.

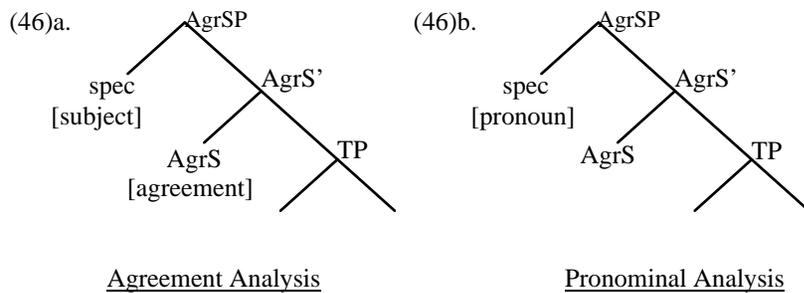
2.5 Subject Agreement versus Subject Pronoun

A current debate in Bantu linguistics centers on whether SA is an agreement marker or a pronominal clitic (e.g., Bresnan & Mchombo, 1987; Keach, 1995; Zwart, 1997)¹⁷. Traditionally, it has been described as agreement (Ashton, 1947; Myachina, 1981), although some of its functions are pronominal in nature. I will discuss the two sides of this debate and determine the nature of SA in Nairobi Swahili. While the importance of this will become clear once we begin discussing the acquisition data, a priori we have reason to pay attention to the nature of SA. We saw in

¹⁷ Similar debates are currently going on for other languages that have rich agreement morphology, e.g., Spanish (Ordoñez, 1997; Goodall, 2002; Grinstead, 1999), Tongan (Otsuka, 2001), etc. Furthermore, there are dialects of English in which preverbal pronouns are analyzed as agreement markers, e.g., Bošković & Chapman (1998) show that in certain non-standard dialects in English, the preverbal pronoun acts more like a subject agreement marker than a pronominal subject, e.g., the ‘pronoun’ is bound to the verb stem, conjunction is impossible, null subjects are licensed, etc. They argue that English pronouns are in the midst of a move from pronouns to agreement clitics (as suggested by Givón, 1975). We will see that Swahili may be in the same process.

chapter 1 that children in other Bantu languages either omit SA entirely or reduce it to a ‘shadow vowel’. We also saw that children acquiring Quechua omit SA at early stages. Thus we may expect Swahili children to have difficulty with SA as well. Therefore a careful analysis of the properties of SA in the adult language is warranted.

Let me begin by clarifying the difference between an ‘agreement analysis’ and a ‘pronominal analysis’. Agreement is a process in which two elements that are in a local configuration (46a)¹⁸ share morphological features through a process of feature matching. Although not crucial to my analysis, I assume that the subject raises from a lower VP position to [spec, AgrSP] (Koopman & Sportiche, 1991) and triggers agreement with the head AgrS. This subject can be optionally null, in which case it is licensed through rich agreement (Rizzi, 1982; Taraldsen, 1978; see section 2.5.1).



¹⁸ Recent work in Minimalism (Chomsky 1998, 2001) has raised the possibility that agreement relations need not be strictly local, as in 38a. I will not consider the possibility of a slightly ‘looser’ agreement configuration because all agreement relations are strictly local in Swahili. In other languages in which non-local relationships occur, we might need to reconsider our understanding of agreement.

A pronoun, on the other hand, is a DP. It is the actual subject raised from [spec, VP] to [spec, AgrSP], and is cliticized to the rest of the verbal complex. The crucial difference between agreement and a pronoun is that agreement is the head of AgrS itself and the subject is in the specifier, while a pronoun on the other hand is a DP in the spec of AgrSP (46b). Under a pronominal analysis, in a sentence with an overt subject, what looks like the subject is in fact a topic that binds the pronominal clitic in subject position. Note, therefore, that under a pronominal analysis a preverbal ‘subject’ in Swahili should have the properties of a topic and not a subject (Zwart, 1997). Distinguishing between these two analyses is not an easy task, as a survey of the Bantu literature reveals.

Bresnan & Mchombo (1987) argue that in Chiche_a, an East Central African Bantu language, SA is ambiguous between agreement and a pronoun. They use several functional diagnostics, including the interplay of word order with agreement morphology and tone with phrase structure. They show that overt subject NPs differ from overt object NPs in their distribution in relative clauses and interrogatives, these being contexts which allow and disallow topics respectively (see the next section for more detailed reasons for why this is so). Chiche_a differs from Swahili in some significant ways,¹⁹ and so it does not benefit us to discuss the details of the

¹⁹ For example, Chiche_a has tone, Swahili does not. Chiche_a does not exhibit the ‘Definiteness Effect’ in Object Agreement (Bresnan & Mchombo, 1987, p.761), while Swahili does. Chiche_a allows postverbal subjects to be questioned, a fact that Bresnan & Mchombo take to mean that ‘the subject and topic NPs appear at the same level of structure in the S, with exactly the same ordering possibilities’ (p.775). Swahili, on the other hand, disallows such questions:

- a. Nani alipiga picha? ‘Who took a picture?’
- b. ??/* Alipiga picha nani? ‘Took a picture, who?’

arguments in Bresnan & Mchombo. However, their methodology for distinguishing agreement from pronouns has been applied to Swahili by Keach (1995).

2.5.1 Keach (1995)

Keach (1995) argues that SA in Swahili, as in Chiche_a, behaves as both agreement and a pronoun. Following Bresnan & Mchombo, she defines agreement as crucially being a local phenomenon (i.e., it always occurs in a spec-head configuration). The process of agreement involves the sharing of \bar{A} -features between the XP (in spec position) and the agreeing marker (the head). A pronominal analysis, on the other hand, is one in which SA is the subject, anaphorically bound by the topic DP. Theta role assignment occurs directly to the SA pronoun, and the theta role is transmitted through a chain to the overt topic DP. Keach presents three data arguments, which yield conflicting results (the first in favor of a pronominal analysis, the other two in favor of an agreement analysis), hence the claim that SA is ambiguous between agreement and a pronoun.

She starts by showing that post-verbal subjects are possible ordinarily, as are subjects raised to the topic position of a higher clause, as in (47b-c, Keach's 4a-c).

- (47)
- a. watu wa Kenya _i wa_i – na – wa – pend – a watoto
 people of Kenya SA_{3pl}-pres-OA_{3pl}-like-IND children
 'People of Kenya like children'
- b. wa_i – na – wa – pend – a watoto watu wa Kenya _i
 SA_{3pl}-pres-OA_{3pl}-like-IND children people of Kenya
 '(They) like children, people of Kenya'

- c. watu wa Kenya _i ni – na – fikir - i kuwa wa_i–na – wa–pend–a watoto
 people of Kenya SA_{1s}-pres-think-IND that SA_{3pl}-pres-OA_{3pl}-like-IND children
 'People of Kenya, I think that, (they) like children.'

According to Keach, these examples are compatible with both a pronominal and an agreement analysis. Under the agreement analysis, agreement occurs before movement, and then the subject DP is moved leftward or rightward as normal. Under a pronominal analysis theta role assignment occurs directly to SA, and is then transmitted through a chain to the overt DP.

She then presents data showing that SA is ungrammatical when the HU- tense marker (indicating habituality) is used:

- (48) a. Watu wa Kenya hu - wa – pend – a watoto
 people of Kenya hab-OA_{3pl}- like - IND children
 'People of Kenya like children'
- b. *Watu wa Kenya wa - hu - wa – pend – a watoto
 people of Kenya SA_{3pl}-hab-OA_{3pl}- like - IND children
 'People of Kenya like children'

In sentences such as (48a), where SA is absent, postverbal subjects and the raising of subject to matrix topic position are ungrammatical as illustrated in examples (49a,b) (cf. 39 b,c):

- (49)a. *Hu – wa – pend – a watoto, watu wa Kenya
 hab-OA_{3pl}- like - IND children people of Kenya
 'like children, people of Kenya'
- b. *Watu wa Kenya ni–na–fikir – I kuwa hu – wa – pend–a watoto
 people of Kenya SA_{1s}-pres-think-IND that hab-OA_{3pl}- like - IND children
 'People of Kenya, I think that, like children'

According to Keach, this is evidence that SA is a pronoun for the following reason: theta role assignment occurs directly to the SA pronoun and is then transmitted through a chain to the topic. Eliminating the SA thus results in a theta-criterion violation. No violation occurs when the topic is local and

can receive its theta role directly (example 40a)²⁰. For Keach, a DP in topic position is sufficiently local for theta assignment to occur, but once that DP is moved (either postposed or raised to a higher clause), that local relationship no longer exists²¹. According to Keach, it is unclear how to

²⁰ Keach provides examples which show that subjects in *Hu*-clauses must be overt:

- i. ulevi hu - ondo - a akili
drunkenness HAB-remove-IND sense
'Drunkenness removes common sense'
- ii. *hu - ondo - a akili
HAB-remove-IND sense

This suggests that the theta criterion must be satisfied through the subject directly in such cases of missing SA. Thomas Hinnebusch informs me that this may not be as clear as Keach suggests. He reports that native speakers do use *hu*-clauses in conversational speech without overt subjects. I will continue to assume Keach's data because my native consultant agrees with these judgments.

²¹ As discussed earlier, according to Keach, in the case of non-habitual clauses the theta role can be transmitted through a chain from the subject pronoun to the topic DP. However, it is not clear why in *hu*-habitual clauses, a theta role cannot be assigned to the trace of the topic and transmitted by the chain that connects it to the moved topic. This cannot be due to the different A/A' status of these chains, as both are A'-chains. Thus I argue the ungrammaticality of postposed topics in *hu* habitual clauses is not due to the status of SA as a pronoun.

The possibility of postverbal subjects is related to the existence of rich agreement. For example, Italian and Spanish are two languages with rich subject verb agreement and that both allow postverbal subjects. On the other hand, English does not have rich agreement and does not allow postverbal subjects. We see this correlation within the same language here: in the presence of SA, postverbal subjects are possible, but not in the absence of SA. An analysis of postverbal subjects goes beyond the scope of this dissertation, but the relevance here is that the ungrammaticality of postverbal subjects in habituales in Swahili does not necessarily entail that SA is pronominal. We will return to *hu*-clauses in 2.10.3.

account for this ungrammaticality under an agreement analysis. We will return to an alternative analysis of *hu* clauses in section 2.10.3 below.

Her second argument is based on three principles proposed by Bresnan & Mchombo (1987) in their analysis of Chiche_a. These principles are stated in (50):

- (50) i. Relative pronouns bear TOPIC function;
- ii. Questioned constituents bear FOCUS function;
- iii. An argument cannot bear both TOP and FOC function in the same clause.

The first of these principles is not relevant for our purposes. Based on the latter two principles, Keach examines subject *wh*-questions in Swahili.

- (51) nani_i a_i - me -end-a ?
who SA_{3s}-pr.prf-go-IND
who has gone?

Swahili does not have *wh*-movement, and so the *wh*-word in (51) is in situ. Under a pronominal analysis, the SA is in subject position and the *wh*-phrase is in topic position. The *wh*-phrase therefore bears TOPIC function. Furthermore, because it is the questioned constituent, by principle (50ii), it bears the FOCUS function as well. Principle (50iii) rules this ungrammatical, as the *wh*-word bears both TOPIC and FOCUS functions. However, as (51) shows, subject *wh*-questions are possible in Swahili. Keach concludes that, assuming the principles in (50), SA cannot be pronominal.

Her final argument comes from idioms. She notes that idiom subjects resist topicalization, as was pointed out by Bresnan & Mchombo. In (52b), the idiom subject *mtindi* 'brew' is topicalized to a higher clause, and this results in ungrammaticality, suggesting the idiom subject is a true subject and not a topic. This again argues in favor of an Agreement analysis.

- (52) a. Ni – li – fikir - i kuwa mtindi u - me – va - a Asha
 SA_{1s}-past-think-IND that brew SA₃-pr.prf.-wear-IND Asha
 ‘I thought that Asha is drunk’
 (lit: I thought that the brew has covered Asha)
- b. *Mtindi ni – li – fikir – i kuwa u - me – va - a Asha
 Brew SA_{1s}-past-think-IND that SA₃-pr.prf.-wear-IND Asha
 (lit: (As for) the brew, I thought that it has covered Asha)

Of the three arguments presented by Keach, two support an agreement analysis and one supports a pronominal analysis. Thus Keach concludes that SA in Swahili is ambiguous between a pronoun and agreement. I will now argue that this is not the case. In fact, the evidence for a pronominal analysis is ambiguous at best, while there is further evidence supporting an agreement analysis. I present this evidence in the next section.

2.5.2 Additional arguments for agreement

In addition to the arguments presented by Keach, there are three further pieces of evidence that I will present that also suggest an agreement analysis. The first two arguments make use of the fact that in a pronominal analysis, SA is the subject and what has traditionally been referred to as subject is actually a topic. One property of topics is that they cannot be quantifiers (Lasnik & Stowell, 1991; Rizzi, 1993):

- (53) a. I did everything
 b. *Everything, I did (it)
- (54) a. Nothing is impossible
 b. *Nothing, (it) is impossible

In Swahili, this restriction also holds. In (55a), the object (*kila kitu*) is in object position, and is ungrammatical when topicalized, as in (55b).

- (55) a. a – li – nunu – a kila kitabu
 SA_{3s} –past–buy–IND every book
 ‘She bought every book’
- b. *kila kitabu, a – li – (ki)–nunu– a [t]
 every thing SA_{3s}–past–(OA₇)–buy–IND [t]
 ‘Every book, she bought’

Thus the restriction on quantified topics holds in Swahili²². Under a pronominal analysis of SA, the preverbal DP is in topic position, and so a quantifier should be ungrammatical. However, as (56) shows, quantifiers are possible in preverbal position, suggesting that the preverbal DP is in subject position.

- (56) a. kila mtoto a – li – nunu– a ki – tabu
 every child SA_{3s}–past–buy–IND 7–book
 ‘Every child bought a book.’
- b. kila ki–tabu ki – li – nunuli – w – a na mtoto
 every 7–book SA₇–past– buy –passive–IND by child
 ‘Every book was bought by a child.’

Furthermore, the answer to a question cannot be a topic:

- (57) a. Who arrived early?
 b. ?? As for John, he arrived early
 c. John arrived early

In Swahili, the preverbal DP can be the answer to a question²³:

- (58) a. nani a – li – fik – a mapema
 who SA_{3s}–past–arrive–IND early
 ‘Who arrived early?’

²³ Thanks to Stan Dubinsky and Ivano Caponigro for discussions on this point.

- b. ?? Juma, a – li – fik – a mapema
 Juma, SA_{3s}–past–arrive–IND early
 ‘Juma, he arrived early.’
- c. Juma a – li – fik – a mapema
 Juma SA_{3s}–past–arrive–IND early
 ‘Juma arrived early.’

When the preverbal DP is topicalized (indicated by ‘comma’ intonation) in example (58b), it is awkward as an answer to the question in (58a). Thus topics cannot be the answer to questions in Swahili. In example (58c), the non-topicalized preverbal DP is grammatical as the answer to the *wh*- question in (58a). This supports the view that the preverbal DP (without ‘comma’ intonation) is not in topic position, but rather in subject position.

Thus a pronominal analysis of SA is implausible. A final argument in favor of an agreement analysis comes from typology. One criterion that distinguishes clitic pronouns from agreement affixes is the freedom of word order: pronouns are generally more free to move relative to the verb, or allow the verb to move around the clitic. For example, Tagalog has a series of clitics, all of which are constrained by a second-position rule (Schachter, 1995, p.1425). The verb can precede the clitic or follow it, as can other words in the sentence, with the only restriction being that the clitic must be in second position. Affixes, on the other hand, must generally remain proximal to the verb, in the same structural configuration, and with the same set of (usually) inflectional elements between it and the verb. For example, languages in the Takic family (a Southern California branch of Uto-Aztecan) have a subject marker that, similar to Swahili, is the focus of debate. Among the languages of the Takic family, SA has been particularly well-studied in four languages: Luiseño, Cupeño, Serrano, and

Cahuilla. In Luiseño, the unmarked word order is shown in (59a) (examples are from Steele, 1995), where the clitic (*up*) is in second position following the subject (*hengeemal*):

- (59) a. hengeemal up heyiq Subject-clitic-verb
 boy 3sg is:digging
 ‘The boy is digging’
- b. heyiq up hengeemal Verb-clitic-subject
 is:digging 3sg boy
 ‘The boy is digging’
- c. * hengeemal heyiq up Subject-verb-clitic
 boy is:digging 3sg

In (59a), the unmarked order is subject-clitic-verb. According to Steele (1995, p.1227), (59b) with the verb preceding the clitic is semantically non-distinct from (59a). (59c) – where the clitic sequence is not second – is ungrammatical. This is also true of two of the other three most well-studied languages: Cupeño and Serrano. Thus the order of the clitic and verb is free, provided the clitic is in second position. However, Cahuilla, has a set of bound pronominal elements that are obligatorily preverbal. Thus the order clitic-verb is grammatical, but verb-clitic is ungrammatical irrespective of whether the clitic is in second position or not. These clitics are “generally taken to be prefixes rather than (pronouns)” (Steele, 1995,

p.1227)²⁴. In making this distinction, Steele (along with Jacobs, 1975; Steele, 1977; Langacker, 1977) uses word order as a diagnostic for whether a subject marker is an agreement affix or a pronominal clitic, with the former being fixed in position with respect to the verb, and the latter being somewhat freer.

We can now apply this test to the Swahili SA marker to determine whether it is a prefix or a pronominal clitic. As described in section 2.1, the verbal complex acts as a unit. When the verb moves, all the preverbal members of the complex move with it. Similarly, when the subject moves, the subject agreement marker remains in its original position, never moving with the subject. In this regard Swahili SA behaves like Cahuilla SA. Word order is fixed with respect to the verb, suggesting that it is an agreement marker rather than a pronominal clitic.

Summarizing, while a pronominal analysis has been proposed by various authors, the evidence that SA is a pronoun is weak and unclear. The evidence that SA is agreement, on the other hand, is considerably stronger. The arguments for this latter position include the fact that idiom subjects behave as subjects and not topics, quantifier DPs may occur in subject position (showing this position to be subject position and not a topic

²⁴ The only examples she gives are to illustrate that these prefixes combine subject and object marking, and not to illustrate the unacceptability of free word order. One example is her example (7a):

‘echem-némiwe
1pl/2sg-chased
‘We chased you.’

The fact that Steele glosses this example with a hyphen between the prefix and verb suggests that it behaves as a single unit, akin to the Swahili verbal complex, and her description of the facts suggests the same.

position), and that the SA prefix behaves like agreement in other languages, as opposed to a pronominal clitic. Thus I conclude that SA is agreement between the subject and the verb. In the next section we will investigate a proposal that tense in Swahili is an auxiliary verb and not an inflectional tense prefix.

2.6 Tense versus Auxiliary Verb

For the same reasons we investigated SA in adult Swahili, we must investigate tense. Children acquiring Bare verb languages or RI languages tend to either omit inflectional affixes or use infinitival morphology, suggesting that tense is somehow a late acquisition. It is important for us to understand the properties of tense in adult Swahili in order to better understand the acquisition data. Traditionally, the tense marker has been described as an inflectional prefix. Recently, this view has been challenged, and the tense marker has been analyzed as an auxiliary verb. Under this view, the verbal complex is composed of the auxiliary (tense) verb with a SA prefix, and the main verb with OA prefix. There are several reasons why tense may be viewed as an auxiliary. I will present the evidence for this position, along with criticisms. First, many tense markers are transparently related to lexical items that are clearly verbal (Zwart, 1997). Examples are shown in (60).²⁵ The argument is that these elements were originally verbs that have developed into auxiliary verbs.

(60)

| <u>Marker</u> | <u>Tense</u> | <u>Origin</u> |
|---------------|-----------------|--|
| -na- | Present | Conjunction, preposition na meaning and/with |
| -ta- | Future | Verb taka meaning ‘want’ |
| -me- | Present perfect | Verb mala meaning ‘complete’ |

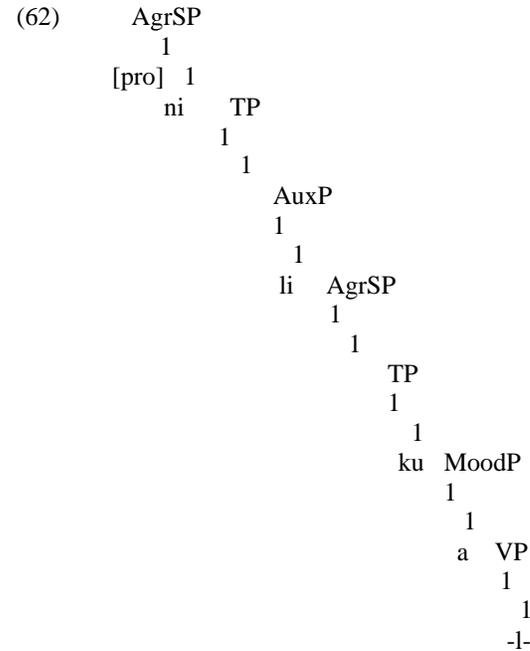
²⁵ The transparency is obvious for *-na-* and *-ta-* but not so for *-me-*.

This claim of diachronic development is legitimate, but not convincing. Givón (1995) argues that SA in Swahili developed from overt pronouns, and is currently in a transition stage between being a pronoun and grammatical agreement. This is supported by the mixed results we saw earlier in our discussion of whether SA is a pronominal clitic or an agreement marker. We concluded that SA in Nairobi Swahili is agreement. Thus while SA may have originated as a pronoun, its current state is that of agreement. Similarly, while tense may have developed from a verbal element, this says nothing about its current state, which must be established from independent examination.

The second argument that T is an auxiliary verb comes from Buell (2000), who shows that in monosyllabic verbs the tense marker is followed by the infinitive marker and the verb, as in (54):

- (61) ni – li – ku – l – a
 SA_{1s}-past-INF-eat-IND
 ‘I ate’

Buell argues that the past tense marker *li* is an auxiliary verb that takes an infinitival complement, hence the infinitive marker *ku*. The structure he argues for is given in (62) below (ignoring unnecessary projections):



I argue against this position. Recall our discussion of monosyllabic verb roots in section 2.4.5. We saw that *ku* (which I call ‘dummy *ku*’) is inserted in these contexts for the purpose of carrying stress, i.e., dummy *ku* is only inserted when the verb stem (the verb root and the mood final vowel) is monosyllabic and a second syllable is required to carry stress (63a-b). When the verb stem is multisyllabic, dummy *ku* is not inserted (63c):

- (63) a. ni – li – ku – l – a
 SA_{1s}-past-INF-eat-IND
 ‘I ate’
- b. ni – li – ku – nyw – a maji
 SA_{1s}-past-INF-drink-IND water

'I drank water'

- c. *ni – li – ku – fik – a nyumba-ni
SA_{1s}-past-INF-arrive-IND house-loc
'I arrived home'

Thus syllabic structure is the crucial determinant in the occurrence of dummy *ku*, which suggests that this is not an infinitive marker, but rather a phonological device used to make the verb stem a well-formed phonological word. Furthermore, despite there being an 'infinitive' marker in these constructions, the interpretation is fully temporal and finite, as indicated by the glosses. Moreover, there is no modal meaning associated with dummy *ku*, as is commonly the case with infinitives in Swahili and other languages (see Stowell, 1981; Duffley 1992; Hyams, 2001). The examples in (64) are taken from Ashton (1947, p.279), showing some cases of Swahili adult root clause infinitives. Both examples carry a modal meaning, as in the examples in (65) from a variety of other languages.

- (64) a. Zama ku – zama we! Swahili
Drown inf-drown you
'Just drown!'
- b. Kwa nini ku–fanya hivi? Swahili
For what inf – do this
'Why do this?'
- (65) a. Niet parkeren hier Dutch
NEG Park-inf here
'No parking here'
- b. What to do?/ Che fare? English/Italian
- c. Non tornare a casa troppo tarde. Italian
Not come-inf. home too late
'Don't come home too late'

The Dutch jussive example in (65a) expresses necessity, expressions such those in (65b) in English and many other languages typically have a modal

meaning, roughly 'what should we do', 'where should we go', and (65c) illustrates that in Italian and many other varieties of Romance negative imperatives are formed with the infinitive (as discussed by Zanuttini 1997 and others). The infinitive morpheme in Swahili too is associated with a modal meaning (examples 64). The fact that dummy *ku* is not modal suggests that it is not a true infinitive.

Finally, dummy *ku* occurs after OA in Nairobi Swahili:

- (66) ni – li – i – ku – l – a
SA_{1s}-past-OA₃-INF-eat-IND
'I ate it'

This is unexpected if *ku* is an infinitive marker in a position associated with (non-finite) tense since TP is higher than AgrOP, and so we would expect OA to follow *ku*. The proximity of the dummy *ku* to the verb root suggests that this is a phonological insertion not related to a syntactic infinitive.

A third argument that tense is an auxiliary verb comes from Buell (2000), who shows that in Standard Swahili sentential conjunction, an infinitive marker occurs in the second conjunct, as in (67):

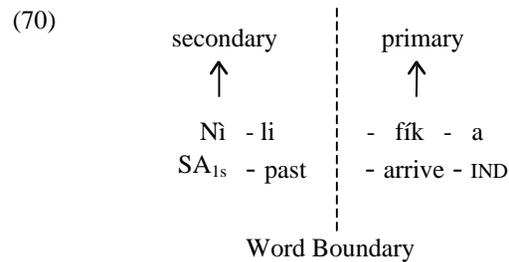
- (67) a – li – kimbi – a na ku–tembe-a
SA_{3s}-past-run-IND and INF-walk-IND
'He ran and walked.'

Such constructions occur only in Standard Swahili, not Nairobi Swahili and so do not bear on this study. In Nairobi Swahili coordination occurs either with the verb stem, as in (68), or with the entire verbal complex, as in (69).

- (68) a – li – kimbi – a na tembe-a
SA_{3s}-past-run - IND and walk-IND
'He ran and walked.'
- (69) a – li – kimbi – a na a – li – tembe – a
SA_{3s}-past-run-IND and SA_{3s}-past-walk-IND
'He ran and he walked.'

Thus conjunction does not provide us with evidence for an auxiliary analysis of tense.

The final argument for tense being analyzed as an auxiliary is the stress pattern exhibited in the verbal complex. As noted in section 2.2, Swahili has a rule of penultimate stress. In a simple verbal complex, this occurs on the main verb. However, secondary stress is also present, usually on the SA marker. Barrett-Keach (1986) interprets the secondary stress as evidence that there is a word boundary between tense and the rest of the verbal complex²⁶. This word boundary may suggest that tense is a separate lexical item, with SA being a prefix to this verb (Zwart, 1997).



In cases where the tense marker is more than a single syllable, the secondary stress shifts rightward, as it would if the right edge of the tense marker is a phonological word boundary.

- (71) a. Nĩ - me - fík - a
 SA_{1s} - perf. - arrive - IND
 'I have arrived.'

²⁶ She doesn't actually argue for an auxiliary analysis of tense, but is more concerned with showing that there is a syntactic constituent Aux, which is manifested in Swahili as SA+T, and is a distinct word.

- b. Nĩ - mèsha - fík - a
 SA_{1s}-perf.comp.-arrive-IND
 'I have already arrived.'
- c. Nĩ - mekwìsha - fík - a
 SA_{1s}-perf.finish-arrive-IND
 'I have finished arriving.'

Secondary stress in (71a) is on the SA marker *ni*, which is the penultimate syllable from the right edge of the tense marker. In (71b), the tense marker is disyllabic, and we see that secondary stress shifts rightwards to the first syllable of the tense marker, remaining on the penultimate syllable from the right edge of the tense marker. In (71c), the tense marker is trisyllabic, and secondary stress is on the medial syllable of the tense marker. Again, this rightward movement of secondary stress results in stress occurring on the penultimate syllable from the right edge of the tense marker. Thus the right edge of the tense marker behaves as a phonological word boundary.

The conclusion that there is a word boundary between T and the rest of the verbal complex is reasonable. However, this is not a convincing argument that tense is an auxiliary verb. If this is a bipartite structure with the tense marker being an auxiliary verb, then it is unclear why the phonological word boundary is not a full-fledged boundary, i.e., it is unclear why stress on the SA would be demoted to secondary status. Furthermore, a word boundary suggests some amount of morphological looseness, in that material may intervene between the words. However, nothing can be inserted into this position. In (72a), the adverb *upesì* 'quickly' occurs utterance finally, as it does in the English gloss. However, unlike the English gloss in (72b), the adverb may not intervene between the 'auxiliary'

and the verb in Swahili. Thus the word boundary that exists between tense and the verb is clearly not loose enough to allow intervening adverbs.²⁷

- (72) a. ni – ta – maliz – a upesi
 SA_{1s}–fut–finish–IND quickly
 ‘I will finish quickly.’
- b. * ni – ta upesi maliz – a
 SA_{1s}–fut. quickly finish–IND
 ‘I will quickly finish.’

The existence of a phonological word boundary is not sufficient evidence to show that tense is an auxiliary as there are numerous other reasons why a boundary may arise (for example, the phonology of Swahili may restrict the length of words, requiring the segmentation of long words into easier-to-pronounce units).

Finally, auxiliary verbs can often be conjoined, e.g., the English example in (73a). However, such conjunction is impossible in Swahili (73b):

- (73) a. I did and he will sing English aux conjunction
 b. *ni – li na a – ta –imb–a Swahili ‘aux’ conjunction
 SA_{1s}–past and SA_{3s}–fut–sing–IND
 ‘I did and he will sing’

In sum, the arguments for tense being an auxiliary are inconclusive. This is one of the areas in which we will look to the acquisition data for further evidence.

²⁷ Furthermore, Zwart (1997) shows that in some relative constructions in which subject inversion is possible, you would expect the subject to intervene between the ‘auxiliary’ and the main verb. However this is impossible, with the only grammatical inverted structure being the post-main-verb construction. Zwart concludes that ‘the auxiliary and the main verb do form a unit of some kind’.

2.7 Object Agreement versus Incorporated Pronoun

In section 2.4.4 we briefly discussed OA. In this section I will present several views of OA, and I will then discuss whether OA is agreement or a pronoun, similar to what we did in section 2.5 for SA. I will outline some arguments put forward by Bresnan & Mchombo (1987) and Keach (1995). I conclude that contrary to Keach’s claim that OA is a pronoun, it is in fact agreement in Nairobi Swahili.

Bresnan & Mchombo (1987), in their discussion of Chiche_a, show that OA is purely an incorporated pronoun. Using tone rules, they show that the object and a non-agreeing verb form a constituent (74), while an object and an agreeing verb do not (75). In the latter case, the verb and the OA marker form a constituent, arguing that the OA is the object. This suggests that the lexical object in an agreeing-verb-structure has been topicalized outside the VP.

- (74) (SA-T) [Ø - V Object]
 (75) (SA-T) [OA-V] [Object]

Secondly, they show that the order Verb-Object is strict when there is no agreement, but when OA occurs, word order is free. This suggests that with OA, the object has incorporated into the verb, and what appears as an overt object is a topicalized object binding the OA pronoun. Swahili differs from Chiche_a in this respect because there is no tonal evidence available in Swahili, and the agreement facts are somewhat different.

Animate objects almost obligatorily require agreement, as Keach (1995) shows:

- (76) a. ni – na – m – pend – a Juma
 SA_{1s}-pres-OA_{3s}-like-IND Juma
 ‘I like Juma’
- b. *ni – na – pend – a Juma
- c. ni – na – ki – som – a ki-tabu
 SA_{1s}-pres-OA₇-read-IND 7-book
 ‘I am reading the book’
- d. ni – na – som – a kitabu
 SA_{1s}-pres-read-IND 7-book
 ‘I am reading a book’

(Keach, 1995, examples 12a-d)

Keach argues that OA is ambiguous between agreement and a pronoun and that animacy is the determining factor: an overt animate object such as that in (76a) can be either a topic binding an incorporated pronoun, or the object agreeing with the verb. However, inanimate objects can never agree with the verb, but rather are always topics binding the incorporated pronoun. In cases where there is no overt OA (76d), agreement has not occurred and the object has not topicalized either. According to Keach, this final option is only possible with inanimate objects.

There are two problems with this argument. First, according to Keach, in every other context of agreement in Swahili, the agreement marker is obligatory²⁸: SA is obligatory with both animate as well as inanimate objects, and OA is obligatory with animate objects. It is unclear why OA is ‘optional’ with inanimate objects. Secondly, the claim that OA is obligatory with animate objects is not correct. Keach fails to mention the

²⁸ Although see section 2.4.1 for exceptions that Keach does not mention, namely the [-SA] clauses.

different interpretations associated with clauses with OA. Notice the difference in translation between (76c) and (76d) above, the two examples with inanimate objects: the latter takes a non-specific reading while the former is specific. Thus OA is associated with a *specific* object. In the animate-object examples, (76a) and (76b), the object in these sentences is a name. Names are obligatorily specific, and so obligatorily take OA (hence the ungrammaticality of 76b). However, if we replace the name *Juma* with a potentially non-specific animate object, the ‘optionality’ returns:

- (77) a. ni – ni – wa_i – pend – a wa-toto_i
 SA_{1s}-pres-OA_{3pl}-like-IND 2-child
 ‘I like the children’
- b. ni – na – pend – a wa – toto
 SA_{1s}-pres-like-IND 2-child
 ‘I like children’

These examples show that while animacy may be a powerful predictor in Swahili, it does not determine OA. Rather, specificity is the determining feature in OA. In fact, we see that OA is never optional, but rather depends on specificity. We shall return to this point shortly.

Another set of data that Keach presents to argue in favor of an ambiguous OA relates to word order. She presents data showing that when OA is present, the direct object can scramble, but in the absence of OA, the position of the direct object is limited. The unmarked word order is given in (78) (Subject-Verbal complex-Object), and each subsequent sentence pair shows that any divergence from that word order requires OA. The order O-S-V is grammatical in (79a) with OA, but ungrammatical in (79b) without OA. The order S-O-V is grammatical in (80a) with OA, but ungrammatical

in (80b) without OA. The order V-S-O is grammatical in (81a) with OA and ungrammatical in (81b) without OA.

- (78) mw-alimu_i a_i – li – wa_j – pend – a wa-toto_j
 1-teacher SA_{3s}-past-OA_{3pl}-like-IND 2-child
 ‘The teacher liked the children’
- (79) a. watoto_j mwalimu_i a_i–li–wa_j–pend-a
 (O) (S) (OA) (V)
- b. *watoto mwalimu a–li–pend–a
 (O) (S) (V)
- (80) a. mwalimu watoto a–li–wa–pend–a
 (S) (O) (OA) (V)
- b. *mwalimu watoto a–li–pend-a
 (S) (O) (V)
- (81) a. a-li-wa-pend-a mwalimu watoto
 (OA) (V) (S) (O)
- b. *a-li-pend-a mwalimu watoto
 (V) (S) (O) (Keach, 1995, examples 14-17)

This data is meant to argue that when scrambling occurs, OA is obligatory. Keach argues that since these are all examples of non-local objects, they must all be topics which can be moved around so long as they are linked to the incorporated pronoun; hence obligatory OA.

I have two objections to this analysis. First, this is a view of agreement as a purely local process, i.e. because OA and the ‘agreeing’ XP surface in a non-local configuration, the process cannot be agreement. However, in our definition of agreement in section 2.5, we saw that

agreement is a process of feature sharing (or checking) between two elements that are in a spec-head configuration *at some point in the derivation*. It is possible for two elements to be in a spec-head configuration at some point in the derivation, only to move into a non-local configuration. Furthermore, more recent views of agreement suggest that it is a loosely local process, i.e., there are instances of non-local agreement (see Chomsky, 1998, 2001 for arguments on the process AGREE). Therefore the data presented in (78)-(81) do not argue for an incorporated pronoun analysis of OA, since the object could just as easily be a topic with agreement.

The second objection arises from the data itself. The examples in (78)-(81) all make use of a past tense marker, indicating that the event that is being denoted is complete and known to the speaker (in its entirety). Therefore, this pragmatically forces a specific reading of *watoto* ‘children’²⁹, hence requiring OA. If we change the tense marker to *na* present tense, which allows for a habitual/generic reading, we can eliminate this bias.

- (82) watoto, mwalimu a – na - pend-a
 children teacher SA_{3s}-pres-like-IND
 ‘Children, the teacher likes them’

This is an extremely difficult judgment to elicit from consultants because of two factors: first, most Swahili speakers have been prescriptively taught that

²⁹ A non-specific, past complete reading would require a very complex context: He used to like children, but then something happened and he doesn’t like children anymore. While this is possible, without this context being explicitly constructed, such an interpretation is extremely unlikely.

animate objects obligatorily require OA, and so they will reject the sentence on that basis³⁰. Speakers who reject the non-inverted version of the sentence (mwaliimu anapenda watoto) because of the lack of OA will also reject the sentence in (82). The examples that Keach provides show that her consultants were such speakers, and so it is not surprising that Keach was led to the conclusions she reached. Second, a topicalized object tends to be previously-mentioned information, and so is usually specific. A non-specific reading such as that in (82) is possible, but requires a very carefully created context.

Keach presents an argument in favor of an agreement analysis for OA. This argument is very similar to that presented in favor of an agreement analysis for SA: idiomatic objects. She argues that if inanimate OA has only a pronominal function (which is her claim), then an idiomatic inanimate object DP should lose its idiomatic interpretation when it occurs with OA:

- (83) a. ni - li- pig-a pasi
 SA_{1s}-past-hit-IND iron
 'I ironed'
 (lit.: 'I hit iron.')
- b. ni - li - i - pig- a pasi
 SA_{1s}-past-OA₃-hit-IND iron
 'I ironed it'
 (lit.: 'I hit it (with) iron.')

The fact that the idiomatic object does not lose its idiomatic interpretation when OA is present argues that OA with inanimate objects is purely

³⁰ I am basing this on my own experiences, as well as reports from all my native speaker consultants.

agreement, not a pronoun.³¹

Summarizing, we have seen that OA is dependent on specificity, and not optional as has been previously thought. We also saw that OA does not affect the interpretation of idiomatic objects, suggesting that OA is agreement, not a pronoun. For these reasons, I will continue to assume that OA is agreement between the object and verb, and that it is triggered by specificity. I assume that it is parallel in all respects to SA, with the exception of the specificity requirement, and so syntactically it should be completely parallel.

I have discussed several issues regarding the three prefixes in Swahili. I will not discuss negation because the children in this study rarely use syntactic negation. For the same reason I will also not discuss the derivational suffixes which occur between the verb and the mood final vowel. For a detailed and thorough analysis of one such suffix, the

³¹ In addition to the idiomatic objects argument, she uses interrogatives to show that inanimate OA is pronominal and animate OA is optionally agreement or pronominal. Her examples are:

- a. a - li - on - a nini? b. *a - li - ki - on - a nini? **Inanimate**
 SA_{3s}-past-see-IND what SA_{3s}-past-OA₇-see-IND what
 'What did he see?'
- c. a - li - mw - on - a nani? d. *a - li - on - a nani? **Animate**
 SA_{3s}-past-OA_{3s}-see-IND what SA_{3s}-past-see-IND what
 'Who did he see?'

These examples show that OA is possible in interrogatives only when the direct object is animate (ex.c-d). This does support Keach's claim that OA is pronominal with inanimate objects but can be agreement with animate objects. However, this data is incompatible with the data presented in the text. Reconciling these facts goes beyond the scope of this project, as my goal here is to show that OA is ambiguous, even if not in the way that others such as Keach have argued.

extensive evidence for verb raising. He shows that when VP ellipsis occurs, the verb itself is not elided, since it has raised out of the VP.

- (85)a. mw-alimu a – li – nunu-a ki-tabu ch-a Chomsky
 1 – teacher SA_{3s}–past-buy–IND 7–book 7–of Chomsky
 ‘The teacher bought Chomsky’s book’
- b. na wa-nafunzi wa – li – nunu – a ~~ki-tabu ch-a Chomsky~~ pia
 and 2–student SA_{3pl}–past-buy–IND ~~7–book 7–of Chomsky~~ too
 ‘And the students did too/bought (it) too,’

I use a strike-through to indicate material that is omitted due to ellipsis. In example (85), ellipsis occurs in the second conjunct (85b). In English when ellipsis occurs, the entire VP is omitted, hence the term VP ellipsis. This is indicated in the English gloss of (85b) (‘And the students did too’), where the verb has been elided, but the presence of ‘do’ indicates that tense remains. However, in Swahili when VP ellipsis occurs, the verb (*walinunua*) remains intact and only the object is omitted (*kitabu cha Chomsky*). This is indicated by the strike-through in the Swahili sentence, and the second English gloss (‘And the students bought (it) too’). Ngonyani argues that the verb raises out of the VP prior to ellipsis, leaving only the object within the VP to be elided³².

³² As an aside, Ngonyani also shows quite convincingly that the applicative structure contains a Larsonian VP shell, with the higher VP headed by the applicative morpheme, and the lower VP headed by the verbal root. Each VP has a separate AGROP. He finds that when VP ellipsis occurs, the verb remains intact (hence, he concludes that V raising has occurred prior to ellipsis). Furthermore, when ellipsis occurs in double object constructions such as the applicative, either both objects are elided or only the lowest object is elided. It is never the case that the left most object is elided with the right most object remaining intact. This shows evidence of a hierarchical structure within the VP.

Demuth & Gruber (1993) and Ngonyani (1996) both propose this structure because of its transparent reflection of the morpheme order in Sesotho and Swahili (i.e. Subject SA – T – OA – Verb Objects). However, assuming a strong version of Baker’s Mirror Principle (Baker, 1985; Baker, 1988), and given the current understanding of incorporation and movement, the morpheme order that is exhibited in Swahili is not immediately derived. Specifically, if we adopt the structure in (84), and assume that all movement is leftward and adjunction is to the left (Kayne, 1994), the order of morphemes that we obtain is V-AGRO-T-AGRS : the exact mirror image of the order that we actually see in Swahili.

Therefore, to account for the correct morpheme order, I propose a series of phrasal movements (as in Deen, 1999). Beginning with the base structure in (84), I follow Ngonyani in assuming that dominating the AGROP projection is a CliticP projection, with OA as its head. I assume a separation of agreement and case: AgrOP assigns accusative case, and OA occurs in a separate projection, which we call CliticP. Assuming that the direct object (DO) must license OA, we assume that the DO raises through [spec, AGROP] checking case features, to [spec, CliticP]. The $_$ -features are checked, and OA is licensed in the syntax³³. The lexical subject, which originates in [spec, VP] raises to [spec, TP] to check case features, and then

³³ The DO actually raises further, to a projection which dominates CliticP. This move has two reasons: first, without such a move, we would have a violation of a principle of syntax which derives from the doubly-filled COMP filter. Namely, the configuration of having an overt head and an overt spec is ungrammatical. The second reason is that the DO is a referential element, and so it moves to the spec of RefP (cf. Beghelli & Stowell, 1992). For details, see Deen (1999).

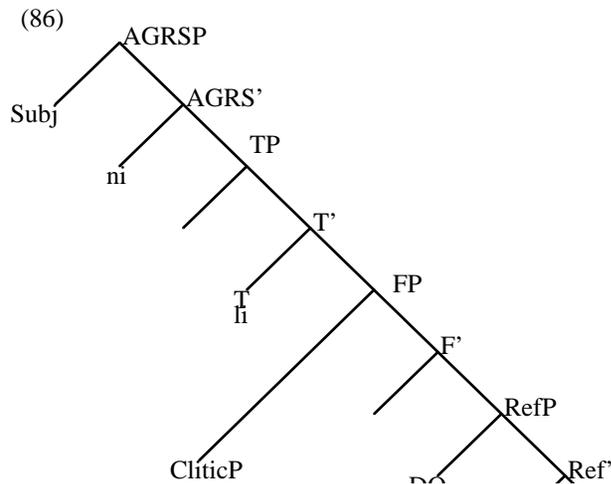
to [spec, AGRSP] to license the SA. I further assume that the verb also raises and left adjoins to the mood marker which is in the head of MoodP (immediately dominating VP). This gives the correct order of V-mood and is consistent with Ngonyani's ellipsis facts.

Given these movements, the morpheme order in the derivation still does not match the observed morpheme order in adult Swahili.

Specifically, the morpheme order at this point is the following:

Subject SA T DO OA V Mood

The DO is too high in the structure. Rather than move the DO rightward (cf. Kayne 1994; Koopman 1996), the cliticP moves leftward and adjoins above DO, as in (86):



There are several points that are worth highlighting about this analysis. First, the analysis involves both phrasal movement as well as head movement. Second, the final phrasal movement is an instance of remnant movement (Webelhuth, 1992; Koopman & Szabolci, 1998). Remnant movement occurs when a phrase which contains the trace of an already moved element moves to a position in which it c-commands the previously moved element. For justification of this type of movement see Koopman & Szabolci (2000) and references therein. Third, there is a branching point after the Tense. This may be related to the stress facts we saw earlier in which the right edge of T acts as a phonological word-boundary. Some theories of the syntax-phonology interface posit that crucial branching

points (or XP boundaries) in the syntactic structure are mapped onto phonological breaks in the phrase (e.g., Selkirk, 1986). This is not incontrovertible evidence for such a position, but is simply suggestive.

And finally, the verbal complex does not constitute a complex head in the syntax. Rather, each morpheme occurs distinctly, with the full morphological verbal complex being created at SPELLOUT (see Julien, 2000 and references therein).

In what follows, I present evidence for various constructions that leads to an analysis of subject agreement in Nairobi Swahili. This remnant movement analysis of Swahili is not crucial to the analysis of agreement omission, but may prove useful in future studies of lower clausal structure of Swahili.

In this chapter so far we have seen that SA is best analyzed as agreement between the subject and the verb, as is OA. I argued that while the evidence for tense being an inflectional clitic rather than an auxiliary verb is less convincing it is nevertheless the best analysis, given the available evidence. We will return to these issues in chapter 4 when we look at child Swahili. In the next section I will discuss null subjects in adult Swahili, and show that the null subject has the characteristics of *pro*, and resembles Italian null subjects in important ways. I will conclude that it is in fact *pro*. In the following section, 2.10, I will introduce a class of clauses that all allow SA omission.

2.9 Null Subjects – null *pro*

Certain languages allow null subjects (e.g., Italian, Spanish) while others do not (e.g., English).

(87) *He said that [e] is eating English

(88) Ha detto che [e] mangia Italian
 have-3s said that eat-3s
 ‘He said that (he) is eating’

In section 2.1, I showed that Swahili allows null arguments. In this section we will investigate null subjects in Swahili. This is relevant to the current study because a hallmark of child language is that subjects may be omitted in obligatory contexts (cf. section 1.2.1). In order to evaluate child language, we must have an understanding of the adult phenomenon to see how children diverge from the adult norm. I will first explain the general theory of null subjects, discussing the licensing condition and the identification requirement on *pro*. I will then show that the null subject in Swahili is *pro*, as in Italian. Swahili null subjects, in addition to satisfying the licensing and identification requirements on *pro*, show other similarities to Italian null subjects.

The fact that null subjects occur in finite clauses in Italian means that the null element is not PRO (as PRO only occurs in non-finite contexts). Furthermore, the null subject in the example above has a definite specific reference, as opposed to an antecedent-controlled reference (as with PRO) or an arbitrary reference (as with PRO_{arb}). Null subjects are thus more akin to overt pronominals. Chomsky (1982) concludes that the null element in subject position in a clause such as (88) is the null counterpart to regular pronouns, and is called *pro*. The distribution of overt pronouns and *pro*, however, are different. For example, in Italian *pro* cannot occur as the object of a preposition, but overt pronouns can:

(89) * Ho parlato con [pro] Italian
 have-1s spoken with

(90) Ho parlato con lui Italian

have-1s spoken with him

It was noted that *pro* occurs in languages that have rich subject-verb agreement such as Italian and Spanish, but not in languages without such agreement such as English or French (Taraldsen, 1978). Furthermore, *pro* occurs in non-subject position in languages that have rich object agreement (such as Pashto, Huang 1989), or rich indirect object agreement (as in Welsh, McCloskey & Hale 1984), as in (91).

- (91) a. ma [*pro*] w_-xwar-a Pashto
 I PRF-eat-OA_{fem-sg}
 'I ate (it-fem)'
- b. *z_ [*pro*] xwr - _m Pashto
 I eat - SA_{1st-masc}
 'I eat (it)'

Both examples in (91) illustrate sentences with a dropped object. (91a) illustrates that in Pashto the object is null in the presence of object agreement on the verb. In (91b), on the other hand, agreement on the verb is with the masculine subject, not with the object, and in this case omission of the object is ungrammatical. Similarly, in Welsh, omission of the indirect object occurs in the presence of agreement between the preposition and the indirect object (example taken from Harbert, 1992):

- (92) Roedd car yn aros amdano [*pro*] Welsh
 was car PRT wait for-Msg
 'A car was waiting for (him)'

Rizzi (1986), capitalizing on these restrictions, concludes that the omission of a pronoun involves rich agreement. However, it has also been noted that some languages with rich agreement do not allow *pro*. For example, German does not allow the omission of referential subjects, but does allow the omission of expletive subjects.

- (93) a. *[e] will zu Hause bleiben German

want at home to-stay
 '(I) want to stay home'

- b. [e] klar ist, daß er nicht kommen wird German
 clear is that he not come will
 '(It) is clear that he will not come.'

These facts have generally been interpreted as pointing to the existence of two distinct conditions on null subjects: a licensing condition, and an identification condition (Rizzi, 1986). The licensing condition applies to all null pronouns, while the identification requirement only applies to referential/argumental null pronouns. The licensing requirement states that a *pro* must be licensed by its governing head. In Minimalist terms this can be interpreted as *pro* having Case features that must be checked. As for identification, in order for a noun to be referential, it must be specified for person / number features. Therefore, the identification requirement states that a referential pronoun must get $_$ -features through co-indexation with a case-governing head. In Minimalist terms, the pronoun must have its $_$ -features checked by an appropriate head. Presumably there is a relation between rich agreement and the existence of $_$ -features on that head in order to allow identification, though the exact specification of "rich agreement" remains elusive. Therefore in Italian *pro* is identified because the language has rich agreement, while in English this is not the case.

- (94) *pro* Parl-o Italiano Italian
 |
 identification

- (95) **pro* speak English English
 |
 no identification possible

How does this solve the problem raised by German? Rizzi claims that German satisfies the licensing requirement, but not the identification

requirement. Thus, German licenses non-referential null pronouns, but because it does not satisfy the identification requirement, null referential pronouns are disallowed.³⁴

How does Swahili fit into this typology of languages? We saw earlier that Swahili has both rich subject-verb agreement, as well as object-verb agreement. We also saw that Swahili allows null subjects and null objects. Therefore, it appears as if Swahili satisfies the identification requirement for null pronouns. The pronouns that are omitted may be referential arguments (expletives do not occur in Swahili), and so I conclude that Swahili satisfies the licensing condition as well. I conclude that Swahili null subjects are *pro*, akin to null subjects in Italian and Spanish (see Khamisi, 1988 for further evidence that *pro* in Swahili occurs in subject, object and indirect object positions). Furthermore, Swahili null subjects have many of the characteristics of *pro* in Italian that differentiate it from PRO. For example, both Swahili null subjects and Italian *pro* alternate with overt DPs:

- (96) *pro* alternates with overt DPs (unlike PRO)
- a. Juma/*pro* a – na – zungumz – a ki – zungu Swahili
 Juma/*pro* SA_{3s}–pres–speak–IND 7–English
 ‘Juma/*pro* speaks English.’
- b. Gianni / *pro* parl–a Inglese Italian
 Gianni / *pro* speak–SA_{3s} English
 ‘Gianni / *pro* speaks English.’

³⁴ Additionally, there are languages such as Chinese that allow subject and object omission without any agreement whatsoever. This is a problem for Rizzi’s proposal. Huang (1984) proposes that these are variables bound by a null topic operator. We shall return to this point in a later section when examining Swahili [-SA] clauses.

Furthermore, in both languages null subjects are possible in matrix, finite clauses (unlike PRO, which only occurs in non-finite clauses):

- (97) *pro* occurs in matrix, finite clauses (unlike PRO)
- a. *pro* a – na – zungumz – a ki – zungu Swahili
 SA_{3s}–pres–speak–IND 7–English
 ‘*pro* speaks English.’
- b. *pro* parl – a Inglese Italian
 speak–SA_{3s} English
 ‘*pro* speaks English.’

Similarly, null subjects in Swahili and *pro* in Italian can both occur in finite embedded clauses (unlike PRO, which can only occur in non-finite embedded clauses):

- (98) *Pro* occurs in embedded finite clauses
- a. ni–na–fikiri [kwamba *pro* a–na–zungumz–a ki–zungu] Swahili
 SA_{1s}–pres–think that SA_{3s}–pres–speak–IND 7–English
 ‘*pro* think [that *pro* speaks English]’
- b. Pens – o [che *pro* parl – a Inglese] Italian
 think–SA_{1s} that speak–SA_{3s} English
 ‘*pro* think [that *pro* speaks English]’
 (See Jaeggli & Safir, 1989 for a full review of *pro*).

In the next section, we will see that SA omission is in fact permissible in more contexts than the habitual clauses discussed in section 2.5.1. I will propose an analysis that involves a null constant (Rizzi, 1992) in subject position, bound by a topic operator, thereby accounting for null subjects in contexts in which the traditional identifier (‘rich agreement’) is absent.

2.10 Subject Agreement Omission

The theory of identification discussed in the previous section predicts that null subjects should not be possible when SA is omitted. However, we will see that null subjects in Swahili are possible in the absence of SA. I will show that there are two sorts of clauses that lack SA: habitual clauses and what I call [-SA] clauses. I will present evidence that the subjects in these two clause types have different properties. The habituals contain true subjects (i.e., in [spec, IP]) while the [-SA] clauses contain a null constant in subject position that is bound by a topic operator. We will make use of this analysis in chapter 5, where we analyze subjects in Swahili child language. We will see that children show knowledge of the properties of null subjects in full clauses as well as [-SA] clauses despite very little overt evidence.

The standard position in the Bantu literature is that the minimal verbal complex in Swahili is as in (99).

(99) SA – T – V – IND

However, to my knowledge there have been no corpus-based studies that examine the question of whether SA may ever be omitted.³⁵ In

³⁵ Carol Meyers-Scotton (p.c.) informs me that SA omission is fairly frequent in the spoken forms of many dialects of Swahili. In fact, she documents such a phenomenon (Scotton, 1969) in the dialects of Baganda and Baluhya speakers in the 1960s. She finds that they frequently omit SA prefixes and rarely use incorrect SA (p. 106). She gives examples such as the following, in which the first line is the dialect, and the second line indicates the Standard Swahili equivalent:

| | |
|--|--|
| a. mi na – sem – a ta–kuw–a dereva | b. li – chez – a m – pila |
| mimi ni–na–sem–a ni–ta–kuw–a dereva | ni–li–chez–a m–pila. |
| I SA _{1s} –pres–say–IND SA _{1s} –fut–be–IND driver | SA _{1s} –past–play–IND 3–ball |
| ‘I am saying that I will be a driver.’ | ‘I played ball.’ |

the dialect of Swahili spoken by the parents of the children in this study, SA omission is noticeably frequent (see below for more details). Furthermore, my native consultant and I agree that SA omission is grammatical, but only under certain conditions. In what follows, I will describe some of these conditions.

In the Swahili literature, four verbal constructions have been described as allowing/requiring the omission of SA:

- the imperative
- Infinitives
- the habitual (marked by *hu*)
- the continuative (marked by *ka*).

2.10.1 Imperatives

(100a) below is an imperative, in which SA is obligatorily absent (note the ungrammaticality of 100b).

- | | | | |
|-------|----|--------------------------------------|----------------------|
| (100) | a. | Pig – a picha! | Imperative |
| | | Hit – IND picture | |
| | | ‘Take a picture!’ | |
| | b. | * U – pig – a picha! | * Imperative with SA |
| | | SA _{2s} – hit – IND picture | |

Cross-linguistically, imperatives are often (although not necessarily) unmarked for subject agreement (see Koopman, 1997). This appears to be a quite general property of imperatives, an explanation of

Nothing in the way of quantitative data is given, and the context for such omission is not indicated. Meyers-Scotton confirms that there have been no corpus based studies to verify this. Duran (1975) also notes that Kipsigi speakers of Swahili allow SA omission (p.76), as well as various non-standard SA markers. However, no quantitative data are provided, so we do not know how prevalent this phenomenon is. This shows that such a phenomenon is not restricted to Nairobi Swahili, but also occurs in other dialects of Swahili.

which goes beyond the scope of this dissertation (see Potsdam, 1995; Platzack & Rosengren, 1997; Zanuttini, 1997; Koopman, 1997 for details). I will discuss the other three types of clauses: infinitives, habituals and continuative clauses, and later extend the analysis of continuative clauses to a general phenomenon of SA omission. I will then describe three significant differences between habituals and continuative clauses. I will use these differences to argue that the syntactic structures associated with these two clause types are significantly different.

2.10.2 Infinitives

Neither overt subjects nor SA can occur with infinitives, as examples (101)-(102) show.

- (101) a. ni – li – jaribu ku–end–a soko – ni Null subject
 SA_{1s} –past–try inf–go–IND market–loc [-SA]
 ‘I tried to go to the market’
- b. * ni – li – jaribu **mimi** ku–end–a soko – ni Overt Subject
 SA_{1s} –past–try **me** inf–go–IND market–loc [-SA]
 ‘I tried to go to the market’
- (102) a. * ni – li – jaribu **ni** – ku–end–a soko – ni Null Subject
 SA_{1s} –pres–try SA_{1s}–inf–go–IND market–loc [+SA]
- b. * ni–li–jaribu **mimi ni**–ku–end–a soko–ni Overt Subject
 SA_{1s}–pres–try **me** SA_{1s}–inf–go–IND market–loc [+SA]

In these examples, as in their English counterparts, the null subject shares the reference of the matrix subject:

- (103) a. John_i tried [e]_i to go to the market Subject Control
 b. * John_i tried [e]_k to go to the market
- (104) a. Juma a_i–li–jaribu [e]_i ku–end–a soko – ni Subject Control
 Juma SA_{3s} –past–try inf–go–IND market–loc

‘Juma tried to go to the market’

- b. *Juma_i a – li – jaribu [e]_k ku–end–a soko – ni
 Juma SA_{3s} –past–try inf–go–IND market–loc

These are subject control verbs, both in English as well as Swahili, and so I assume that the null element in Swahili is PRO, as it is in English.

Additionally, PRO occurs with object control verbs and in arbitrary contexts:

- (105) Mariam_i a – li – mw–omb–a Juma_k [e]_{*i/k} ku–lal–a chini Object Control
 Mariam SA_{3s}–past–OA_{3s}–ask–IND Juma inf–sleep–IND down
 ‘Mariam_i asked Juma_k PRO_{*i/k} to sleep down (on the floor).’
- (106) Ku – ondok – a mapema si mzuri Arbitrary PRO
 inf – depart – IND early not good
 ‘To leave early is not good’

Thus I conclude that PRO occurs in subject position in infinitives in Swahili, as it does cross-linguistically.

2.10.3 Habituals

Turning now to habituals, recall examples (48, taken from Keach, 1995), repeated here as (107). We see that subject agreement is obligatorily absent in habitual constructions (cf. 107b):

- (107) a. wa – tu wa Kenya hu – wa – pend–a wa – toto
 2-person of Kenya HAB–OA₂–like–IND 2-child
 ‘People of Kenya like children’
- b. * wa–tu wa Kenya **wa** – hu – wa – pend–a wa – toto
 2-person of Kenya SA₂–HAB–OA₂–like–IND 2-child

Furthermore, as Keach (1995) reports, the subject in a habitual clause is obligatorily overt:

- (108) a. ulevi hu – ondo – a akili

drunkenness HAB–remove–IND sense
 ‘drunkenness removes common sense’

- b. * hu – ondo – a akili
 HAB–remove–IND sense

We will return to an analysis of hu- habituals shortly.

2.10.4 Continuative Clauses

The continuative construction is a regular ‘tensed’ clause that is used in narratives. It signals a continuation in the time line and is marked with the *ka* morpheme. *Ka* occurs in the same position that tense occurs, in complementary distribution with other tense markers, and so is considered a regular tense in the traditional Swahili literature (see section 2.4.2).

- (109) a. a – ka – kimbi – a na – o
 SA_{3s}–cont–run–IND with–rel.
 ‘(And then) he ran off with them.’
- b. * a – li – ka – kimbi – a
 SA_{3s}–past–cont–run–IND
- c. * a – ka – li – kimbi – a
 SA_{3s}–cont–past–run–IND

A continuative clause usually takes SA like other tensed clauses, as in (110a) below. However, Ashton (1947) notes that the SA marker may be omitted in certain contexts (cf. 110b, where I have used Ø to indicate that SA has been omitted). She describes the resulting interpretation as expressing ‘some emotional quality like mild surprise’ (p.134):

- (110) a. a – li – ib – a wa–toto a – ka – kimbi – a na – o
 SA_{3s}–past–steal–IND 2–child SA_{3s}–cont–run–IND with–rel.
 ‘He stole the children and he ran off with them.’

- b. a – li – ib – a wa–toto Ø ka – kimbi – a na – o
 SA_{3s}–past–steal–IND 2–child Ø cont–run–IND with–rel.
 ‘He stole the children and actually ran off with them.’

2.10.5 Differences between Habituals and Continuatives

These two constructions differ in several important respects, a few of which we have already seen. I will describe three differences in these two clause types: optionality of SA, optionality of subjects, and embedding. I will argue that the omission of SA in habituals results in the lack of an identifier and hence null subjects are prohibited. I will also show that continuative clauses allow null subjects in the absence of SA, which is unexpected given our theory of identification (see section 2.9 earlier). I will then show that continuative clauses are part of a broader class of clauses called [-SA] clauses. These clauses are part of colloquial spoken Swahili and are very frequent in child Swahili.

2.10.5.1 Optionality of SA

First, SA in the habitual clause is obligatorily null (as the ungrammaticality of example 109b shows), while the SA in the continuative can be overt (as in example 110a above) or null (as in example 110b above). This null option is pragmatically marked, but in the appropriate contexts, completely grammatical. Judgments on these facts are extremely clear.

2.10.5.2 Optionality of subjects

Second, the subject of the habitual clause must be overt:

- (111) a. ulevi hu – ondo – a akili
 drunkenness HAB–remove–IND sense
 ‘drunkenness removes common sense’

- b. * hu – ondo – a akili (Examples cited in Keach, 1995)
 HAB–remove–IND sense

However, the subject of a continuative may be either null or overt. In the unmarked case (when the subject of the continuative clause is the same as the subject of the previous discourse), the subject is null. However, the subject may be overt when there is a change in subject or a clarification required. For example, in (112a), the subject of the second (continuative) clause is the same as the subject of the main clause. Similarly, in (112b) the subject of the second clause must be the same as the subject of the first clause if the subject is null. When the subject of the second clause is not identical to the subject of the first clause, as in (112c), then an overt subject is required³⁶.

(112)

- a. Juma a – li – fik – a nyumba–ni Ø – ka – lal – a.
 Juma SA_{3s}–past–arrive–IND home–loc Ø–cont–sleep–IND
 ‘Juma arrived home and (he/*she/*they) then actually went to sleep.’
- b. Juma na Mariam wa – li – fik – a nyumba–ni Ø – ka – lal – a.
 J. and M. SA_{3pl}–past–arrive–IND home–loc Ø–cont–sleep–IND
 ‘J. and M. arrived home and (they/*he/*she) then actually went to sleep.’
- c. Juma na Mariam wa – li – fik – a nyumba–ni. Juma Ø – ka – lal – a.
 J. and M. SA_{3pl}–past–arrive–IND home–loc Juma Ø–cont–sleep–IND
 ‘Juma and Mariam arrived home. Juma then went to sleep.’

So, the subject in continuative clauses may be null or overt depending on discourse considerations, while the subject in habitual constructions must be overt.

³⁶ While our eventual goal is to understand SA omission, I am exemplifying here the fact that subjects may be overt or null in continuative clauses in general. This same fact is true in continuative clauses that are missing SA.

2.10.5.3 Embedding

A habitual clause can occur in an embedded context as in (113), while [-SA] continuative clauses cannot, as shown by the contrast in (114a):

- (113) a – li – ni – ambi–a [kwamba wa – tu wa Kenya
 SA_{3s}–past–OA_{1s}–tell–IND that 2-person of Kenya
 hu – wa – pend–a wa – toto]
 HAB–OA₂–like–IND 2–child

‘He told me [that people of Kenya like children]’

- (114) a. a – li – ni – ambi–a [kwamba a – ka – kimbi – a]
 SA_{3s}–past–OA_{1s}–tell–IND that SA_{3s}–cont–run – IND
 ‘He told me that he then ran off’
- b. ?? a – li – ni – ambi–a [kwamba Ø ka – kimbi – a]³⁷
 SA_{3s}–past–OA_{1s}–tell–IND that Ø cont – run – IND
 ‘He told me that (he) then ran off’

The differences that we have seen so far are summarized in table

2.12:

Table 2.12 Summary of characteristics of habitual and continuative clauses

| | SA | Overt Subject | Can be Embedded |
|--------------|----------|---------------|-----------------|
| Habituals | * | Required | Yes |
| Continuative | Optional | Optional | No |

Recall from the discussion in 2.9 that null subjects must be identified (Rizzi, 1982; Jaeggli & Safir, 1989). Identification can occur

³⁷ My consultant considers this sentence ungrammatical. My judgment is somewhat less clear, but certainly degraded.

through several means, such as control (in the case of PRO), the presence of a c-commanding antecedent (in the case of a trace), or through rich agreement features on a licensing head. This rich agreement provides an identifier for null *pro*. We saw earlier in this chapter that the null subject of a Swahili tensed clause is *pro*. However, in the case of habitual and [-SA] continuative clauses, this rich agreement is missing. Therefore the question arises as to what the status of the null subject is in such clauses.

Specifically, given that the null subject occurs in the absence of SA, how is the identification requirement satisfied?

We saw in the examples in (111) and the summary in table 2.12 that habituals simply do not allow null subjects. Therefore, the answer to the question for habituals is clear: because rich agreement is absent, null subjects are blocked. This is consistent with our theory of identification of null *pro*³⁸. I therefore assume that subjects in habitual constructions are structurally in subject position and must be overt because of the lack of an identifier. However, continuative clauses allow the omission of SA, and in those same clauses a null subject is possible. This is not expected under our theory of identification. Additionally, the fact that [-SA] continuative

³⁸ This is an alternative view of the facts presented by Keach (1995) in section 2.5.1. Recall that habituals do not allow postverbal subjects. Keach argues this is because SA is a pronoun, and in the absence of this pronoun or a local (preverbal) subject, $_$ -role assignment is impossible in habitual clauses. This may well be true, but another view is that the presence of rich agreement allows a free word order. In Italian (a language with rich agreement), postverbal subjects are allowed. However, in English (a language with no rich agreement), postverbal subjects are not allowed. Therefore agreement seems to be important in allowing postverbal subjects, not necessarily the existence of a clitic pronominal subject, as Keach argues.

clauses are not possible in embedded contexts is surprising (cf. examples 114), as embedded *pro* clauses are possible in Italian, as well as in Swahili full clauses:

- (115) *pro* so che cosa *pro* hai detto Italian
 know-1sts what thing have-2nds said
 ‘(I) know what (you) said’
- (116) *pro* ni – na – ju – a *pro* u–li–sem–a nini Swahili
 SA_{1s}–pres–know–IND SA_{2s}–past–say–IND what Full Clause
 ‘(I) know what (you) said.’

Thus, while *pro* is attested in Swahili full clauses, we have evidence that the null subject in [-SA] continuative clauses is an empty category of a different sort. We will see that SA omission is extremely frequent in child language, and thus it is important to understand what kinds of empty categories occur in the input language. Thus we will investigate the omission of SA and the properties of the null subject in such clauses.

Unfortunately the descriptive evidence available in the literature as to when SA may be omitted is very limited. In order to gain a better empirical understanding of SA omission, I investigated the use of SA by the adults in the Swahili corpus. The first thing I looked at were habitual clauses: not a single utterance containing the *hu-* prefix occurred in the entire corpus. Secondly, there were also no cases of continuative *ka*. This is most likely due to the context of the recordings. The continuative *ka* is used to tell stories, and the purpose of these recordings was to elicit stories (or any speech) from the children. Thus, continuative *ka* never occurred in the recordings. However, I discovered that SA was omitted in a significant proportion of adult speech in non-continuative contexts. In the next section, I will discuss the contexts of these [-SA] clauses in Nairobi Swahili. I will then provide an analysis of these clauses which postulates a null constant

(Rizzi, 1992; 1997) as the null element in subject position. Later in chapter 5 I will extend this analysis to the speech of children, showing that this analysis sheds light on other underspecified forms in child speech.

2.11 [-SA] Clauses

I conducted a CLAN analysis on the Swahili corpus targeting the adult utterances in 16 files sampled from all four children. I investigated the omission of SA, the expression of tense in these clauses, the identity of the missing referent, and the occurrence of overt subjects in these clauses. Most of the examples that I will provide come from the spontaneous speech of the adult speakers in the Swahili corpus. However, all examples have been verified with my native consultant (as well as my own judgments), and differences in judgments are noted.

2.11.1 Frequency of [-SA] Clauses

Of the 1470 indicative verbal clauses coded for the adults, 72 (4.9%) are missing SA. Other underspecified clauses (clauses missing tense and clauses missing both tense and SA) account for a combined 1% of indicative clauses. The remaining 94% of indicative clauses are full clauses.

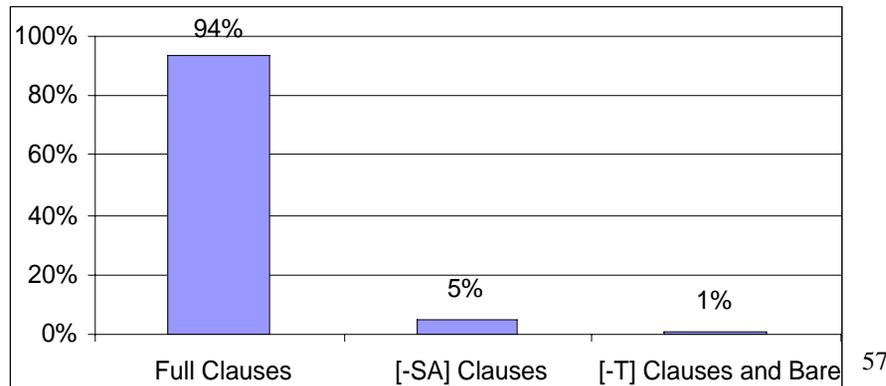


Table 2.13 Proportions of different clause types in adult Swahili.

| Full Clauses | [-SA] clauses | [-T] clauses | Bare Stems | Total |
|--------------|---------------|--------------|------------|-------|
| 1380 (93.9%) | 72 (4.9%) | 14 (0.9%) | 4 (0.3%) | 1470 |

2.11.2 Tense in [-SA] Clauses

[-SA] clauses occur with a variety of tense markers³⁹:

- (117) Ø na – tak – a ch–ai? Present tense
pres–want–IND 7–tea
'(Do you) want tea?'
(Hamisi, HAW05)
- (118) Ø ta – ku – chun – a Future tense
fut–OA_{2s} – pinch–IND
'(I) will pinch you'
(Mot, MUS10)
- (119) ile ni nini Ø me – lal – a pa – le ? Pr. Perfect
that is what pr.perf – sleep – IND loc – there
'What is that that has slept over there?'
(lit: that is what has slept there?)
(Joki, HAW01)

2.11.3 Implicit Reference of [-SA] Clauses

In [-SA] clauses in Swahili, there is no restriction on the implicit reference of the subject. Dropped SA markers can refer to 1st, 2nd and 3rd person referents:

- (120) Ø ta – ku – chapa – a 1st singular
fut–OA_{2s} – slap–IND
'(I) will slap you'
(Sam, MUS10)

³⁹ None of the adults used the past tense marker in [-SA] clauses in this corpus. However, my consultant considers the past tense in a [-SA] clause grammatical, and in child speech the past tense marker is used on several occasions.

(121) Ø na – ju – a ku–wach– a kelele ? 2nd singular
 pres–know–IND inf–leave–IND noise (Ala, MUS09)
 ‘(Do you) know how to stop making noise?’

(120) n–dege Ø na – ruk – a 3rd singular
 9–bird pres–fly up–IND (Ali, FAU07)
 ‘The bird is flying up’

However, there is an asymmetry between [-SA] clauses that have null subjects and [-SA] clauses that have overt subjects. Of the 43 [-SA] clauses that occur with a null subject, the reference of 39 could be determined from context, of which all 39 referred to 1st or 2nd person. Of the clauses that had an overt subject, the subjects were a mix of all three persons. We will return to this point at the end of the chapter.

2.11.4 Overt Subjects in [-SA] Clauses

I conducted a CLAN count of the subjects in full clauses and [-SA] clauses in the Swahili corpus. The results show that in full clauses, adults use overt subjects 16.7% of the time (230 out of 1380) while in [-SA] clauses, adults use overt subjects 40% of the time (29 out of 72).

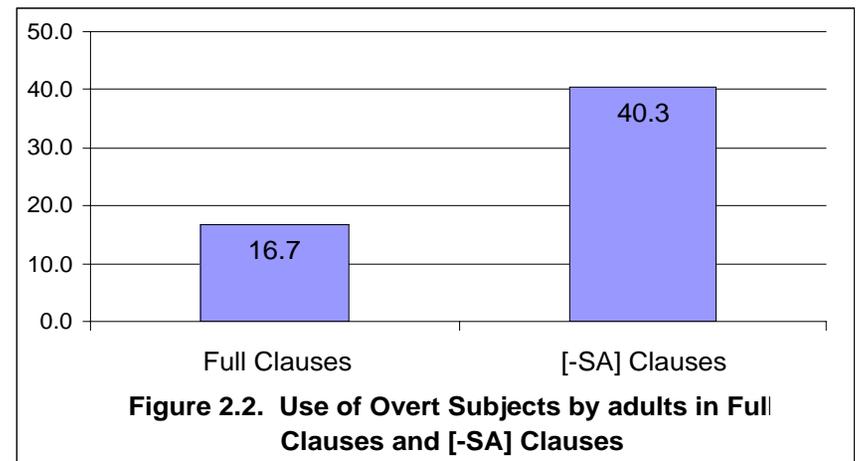
Our theory of identification predicts that null subjects should be completely absent in [-SA] clauses because of the absence of an identifier. Therefore the fact that subjects do not increase to nearly 100% is surprising. In fact, null subjects are still the predominant form in [-SA] clauses – a fact that our theory of identification cannot account for. Below are examples of [-SA] clauses with overt subjects as well as with null subjects (the Ø indicates the missing SA):

(123) a. wewe Ø ta–kul – a ch–akula? Overt Subject
 You fut–eat–IND 7–food (Ala, MUS08, line 230)
 ‘Will you eat food?’

b. ndege Ø na – ruk – a Overt Subject
 bird pres–climb–IND (Ala, MUS12, line 2372)
 ‘The bird is climbing.’

(124) a. ndio, Ø ta – i – beb – a Null Subject
 yes fut–OA–carry–IND (Ali, FAU01, line 178)
 ‘Yes, (I) will carry it.’

b. Ø na – tak – a ice Null Subject
 pres – want– IND ice (Ham, HAW05, line 135)
 ‘Do (you) want ice?’



In the next section we will investigate how null subjects can occur in [-SA] clauses, given that agreement is generally seen as necessary to identify null *pro*.

2.12 Null Subjects in [-SA] Clauses

In the last section we saw that Swahili has a class of clauses in which a null subject appears without an identifier. The primary characteristics of these [-SA] clauses are given in (125):

- (125) a. SA omission is optional (2.10.5.1, figure 2.2).
 b. Overt subjects alternate with null subjects (2.10.5.2).
 c. They cannot occur in embedded context (2.10.5.3).

These clauses occur relatively infrequently when compared to full clauses, but when they do occur, they occur primarily with null subjects. We conclude that the null subject is not *pro*, but some other null element that receives identification through some means other than agreement. This null element is syntactically active, as seen in the following [-SA] examples. In (126), the null subject is the antecedent to the reflexive prefix *-ji-*. In (127), the null subject is the controller for the embedded PRO⁴⁰.

- (126) Ø na – ji – on – a
 pres – refl. – see – IND
 ‘(I) see myself.’

- (127) Ø_i na – ju – a PRO_i ku – onge – a?
 pres – know – IND inf – speak – IND
 ‘Do (you) know how to speak?’

The inventory of null elements permitted by UG includes: *pro*, PRO, NP-trace, wh-trace. We have already seen that *pro* cannot be the subject for [-SA] clauses, so we will now consider whether any of the other null elements are possible subjects for [-SA] clauses. By process of elimination, I will show that none of these null elements satisfy the properties in (125). I will then argue that the null element is a null constant (Rizzi, 1992) bound by a topic operator.

Let us begin by discussing PRO. PRO is the null element that occurs in the subject position of certain non-finite clauses:

⁴⁰ The reflexive example is a constructed example, and the PRO example is an actual utterance from the Swahili corpus (MUS09, line 131).

- (128) a. I entered the race [PRO feeling strong and confident]
 b. PRO to win the race is important.
 c. John tried [PRO to win the race]

We saw earlier that PRO occurs in non-finite clauses in Swahili. However, we can rule PRO out from [-SA] clauses for three reasons. First, PRO occurs prototypically in embedded clauses, and as we saw in (125c), [-SA] clauses do not occur as embedded clauses. Second, PRO occurs in tenseless clauses, while [-SA] clauses always occur with Tense (cf. Examples 117-119). Third, PRO does not usually alternate with overt DPs:

- (129) a. I entered the race [PRO/*Me feeling strong and confident]
 b. PRO/*John to win the race is important
 c. John tried [PRO/*John to win the race]

We saw that in [-SA] clauses, subjects can be overt or null. This is unexpected if the subject is PRO.

Next, let us consider NP-trace. We can eliminate an NP trace from consideration because NP-traces do not alternate with overt DPs either:

- (130) a. John_i seems [t_i to have left]
 b. *John seems [he to have left]

Furthermore, NP-traces must be antecedent-bound in order to fulfill the ECP:

- (131) a. John_i, I like [t_i]
 b. *I like [t]

We saw earlier that approximately 60% of [-SA] clauses have a null subject with no overt preverbal DP. Therefore, if the null subject is an NP trace, 60% of [-SA] clauses do not contain an antecedent DP that could bind the NP trace:

- (132) [t] ta –end–a koti–ni
 _____ fut–go–IND koti–loc
 no antecedent ‘(I) will go to court’

This violates the ECP, and should result in ungrammaticality, contrary to fact. Therefore the null subject cannot be an NP-trace.

A wh-trace has the properties of a variable (Lasnik & Stowell, 1991; Haegemann, 2000) If the null element in subject position in a [-SA] clause is like a wh-trace, it should have the properties of a variable, for example it can be bound by a quantificational element. We see that in [-SA] clauses quantified antecedents are either ungrammatical or marginal at best⁴¹:

- (133) a. Kila mw-anafunzi a – na – som – a ki – tabu
 Every 1-student SA_{3s}–pres–read–IND 7–book
 ‘Every student is reading a book.’
- b. * Kila mw-anafunzi na – som – a ki – tabu
 Every 1-student pres–read–IND 7–book
- (134) a. Wa–tu w–ote wa – na – pig – a kelele
 2-person 2-all SA_{3pl} – pres – hit – IND noise
 ‘Everyone is making noise’
- b. ??/* Wa – tu w–ote na – pig – a kelele
 2-person 2-all pres–hit–IND noise

This suggests that the null element in subject position is NOT a variable, and thus cannot be a wh-trace.

Summarizing, we have found that the null element in subject position of a [-SA] clause cannot be *pro* (no identifier), it cannot be PRO (doesn’t occur in embedded clauses), it cannot be an NP trace (NP traces do

⁴¹ There is variation in judgments on this point, as my consultant disallows all quantificational antecedents to [-SA] clauses, but I find wh- antecedents marginal and other quantifiers ungrammatical. An additional test for a variable is whether it is sensitive to weak crossover effects (Lasnik & Stowell, 1991). Swahili does not have wh- movement, and so this is difficult to test.

not alternate with overt DPs), and it cannot be a wh- trace (it cannot be bound by a quantified antecedent).

2.13 Rizzi’s Null Constant

Rizzi (1992), following Lasnik & Stowell (1991), proposes a new type of null element: a *null constant*. He defines a null constant as:

- a definite description
- <-anaphoric, -pronominal>
- a non-variable
- an R-expression

While overt definite descriptions are free to pick up their referent from the discourse, the null version is subject to the identification requirement that all null elements are subject to. He distinguishes the null constant from a null variable. A null variable must be chain connected to a true quantifier for identification, while a null constant (which is -variable) must be chain connected to a non-quantifier (because the Bijection Principle (Koopman & Sportiche 1982; Chomsky 1986) bars vacuous quantification). Thus, a null constant cannot be assimilated to a wh-trace, for example. According to Rizzi, this non-quantifier is typically a null anaphoric operator in an A’-position. The operator cannot be in an A- position because the null constant is an R-expression and thus cannot be A-bound. Thus, the structure proposed by Rizzi is as follows:

- (135) [_{TOPP} OP [_{IP} nc [_{VP} ...]]]
 |_____|

This binding relation allows identification of the null constant, while also providing a link into the sentence for the anaphoric topic operator.

This structure holds for modern colloquial German in which it is possible to drop a main clause subject in a V2 construction, i.e., from spec-CP. Rizzi shows that while this is possible in main clauses, the possibility disappears in embedded clauses or when the COMP position is filled, whether V2 has applied or not (Rizzi's examples 14):

- (136) a. *(Ich) habe es gestern gekauft* German
 '(I) have it yesterday bought'
- b. *Wann hat *(er) angerufen?*
 'When has he telephoned?'
- c. *Hans glaubt *(ich) habe es gestern gekauft.*
 'Hans believes I have it yesterday bought.'
- d. *Hans glaubt da_ *(ich) es gestern gekauft habe.*
 'Hans believes that I it yesterday bought have.'

Interestingly, the omission of arguments extends to objects as well:

- (137) *(Das) habe ich gestern gekauft.* German
 'This have I yesterday bought.'

Rizzi notes that this has led researchers to conclude that the examples in (136) and (137) involve topic drop (Ross, 1982), as shown in the structures below.

- (138) a. [_{CP} OP *habe* [_{IP} nc *es gestern gekauft*]] (=136a)
 b. [_{CP} OP *habe* [_{IP} *ich* nc *gestern gekauft*]] (=137)

However, Rizzi notes a fact first pointed out by Cardinaletti (1991), that in colloquial German there is an asymmetry between subject drop and object drop. Cardinaletti claims that 'subject drop can involve pronouns of any specification, while object drop is restricted to 3rd person'. She claims that 3rd person specification is a property inherent to operators,

and concludes that subjects should not include an operator. Rizzi therefore limits the above structure (135) to object omission, and proposes that the structure for German subject omission is as follows:

- (139) [_{CP} nc *habe* [_{IP} t *es gestern gekauft*]]

Thus, the null constant is in the specifier of CP, binding an NP-trace in spec-IP position. Since this structure involves no operator at all, the limitation to 3rd person is removed.

Assuming this structure in (139), Rizzi now must explain how a null element (the null constant) can occur in the structure in violation of the identification requirement. He claims that the identification requirement is basically the ECP, stated below:

- (140) Empty Categories <-P> must be chain-connected to an antecedent.

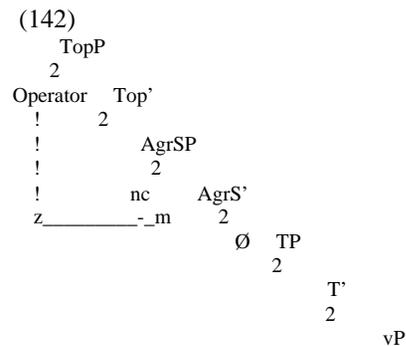
The structure in (139) violates the ECP as stated above, but Rizzi proposes an addition to the ECP:

- (141) Empty Categories <-P> must be chain-connected to an antecedent... if they can be.

He invokes a notion of the "privilege of the root", whereby elements that ordinarily require binding are exempt from this requirement because they are in the root clause in a position that cannot be commanded. Therefore the null constant in structure (139) is exempt from the identification requirement because it is in the specifier of the root and thus *cannot* be clause-internally identified. He suggests that in this case identification occurs through discourse.

2.14 Null Constants in Swahili

I will adopt Rizzi's proposal for Swahili, and show that [-SA] clauses involve a null constant bound by a topic operator. However, I will diverge from Rizzi's analysis with respect to the question of reference. Instead I will claim that the restriction of object drop to 3rd person referents in German comes through discourse identification restrictions (Gutman, 1999), not syntactic restrictions on the discourse operator. I will adopt the following structure:



Assuming a structure as in (142) for Swahili [-SA] clauses, there are various issues that must be resolved. First, we must account for the various characteristics of [-SA] clauses. These characteristics are summarized in (143):

- (143) [-SA] clauses have the following characteristics:
- a. Cannot occur in embedded context
 - b. Subject can be overt or null
 - c. Can occur with all tenses
 - d. Subject cannot be a quantifier

Additionally, we must resolve the question of what identifies the null constant in the absence of SA or any c-commanding antecedent.

2.14.1 Accounting for [-SA] Characteristics

Let us first consider how this structure can account for the characteristics of [-SA] clauses given in (143a-d), returning in section 2.14.2 to the question of identification. The first characteristic is that [-SA] clauses never occur in embedded clauses. Under earlier theories of the left-periphery, this result could be derived through the fact that the operator occupies the spec-CP position, and so is in complementary distribution with complementizers. However, under Rizzi's (1997) articulated left periphery hypothesis, this is no longer tenable. Instead, I propose that the restriction to root clauses is due to the nature of the topic operator. The operator is an anaphoric topic operator, and thus looks to discourse for a topic antecedent. If embedded as a syntactic complement, it does not have direct access to discourse, and so cannot occur in such a configuration. It must therefore be in the specifier of the root. Evidence for this comes from the fact that the operator is optionally null⁴² – a typical characteristic of the root (Rizzi, 1997).

This leads to the second characteristic: the subject can be either overt or null. When we refer to the 'subject', we are referring to the preverbal DP, which in this case is the anaphoric topic operator. Rizzi's (1997) description of this construction is as follows:

...the licensing of null constants is not freely available, but is restricted to a designated kind of A'-binder, the anaphoric operator (an element inherently characterized as an operator but different from quantificational operators in that it does not assign a range to its bindee; rather, the anaphoric operator seeks for an antecedent, to which it connects its

⁴² The question of *when* the operator can be overt or null is left open at this point. Presumably this turns on discourse conditions, as Rizzi suggests. A clearer understanding of what these conditions are is obviously important, but I must leave it for future study.

bindee); anaphoric operators are typically but not necessarily null.

Rizzi, 1997; p.293

Rizzi thus describes a three-member chain (discourse antecedent – anaphoric operator – null constant) in which the anaphoric operator can be optionally null or overt. He describes this as a parametric distinction that some languages allow and others do not, but I propose that Swahili allows both options.

Third, [-SA] clauses can occur with all tenses. This is unsurprising in the structure in (142) as the exact specification of tense is irrelevant to anything in the structure.

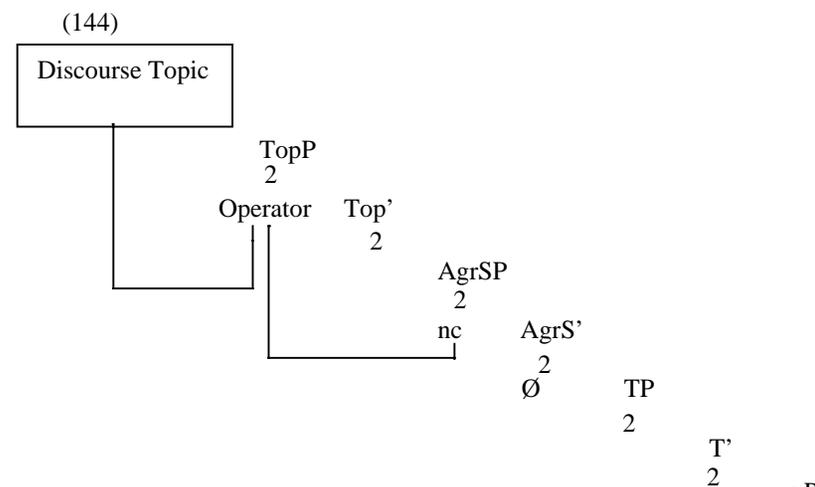
The fourth characteristic is that the subject cannot be quantificational. The anaphoric operator is different from a quantificational operator, in that it ‘does not assign a range to its bindee’. Therefore the anaphoric operator cannot be quantificational, and since it is the anaphoric operator that surfaces as a preverbal DP, it follows that the subject is not quantificational. It is a property of topics in general that quantification is disallowed (Rizzi, 1997), and so it follows that since the operator is in topic position, quantification should not be possible. We have thus accounted for the four characteristics of [-SA] clauses with the proposal that they contain an anaphoric topic operator that licenses a null constant in subject position.

2.14.2 Identification in [-SA] Clauses

Let us now consider the identification requirement. In the configuration in (142) above, the null constant has no SA to check its $_$ -features, and thus is not identified through this kind of feature checking. The only other possibility is identification through the Operator in spec-TopP position. However, we are now faced with a conflict with Rizzi’s

claim (from Cardinaletti, 1991) that operators are intrinsically restricted to 3rd person reference. We saw earlier that reference in [-SA] clauses is not restricted to 3rd person subjects, but is free to refer to all persons. How can we resolve this conflict?

I suggest that the answer lies in the nature of the operator. I suggest that the operator is of the following sort: as Rizzi himself alludes, it is an anaphoric topic operator. The purpose of an anaphoric topic operator is to provide a link for the discourse topic into the internal structure of the sentence. Therefore a topic operator links the reference ($_$ -features) of the discourse topic to its bindee within the sentence (cf. Huang’s 1984 proposal for null arguments in Chinese, a ‘discourse-oriented’ language). Indeed all topics require this link into the sentence, whether the link is through a trace or other means. In this case, the link is through the binding relation with the null constant. The topic operator receives its $_$ -specification from the discourse, and then through a process of feature matching, checks the feature specification on the null constant. This provides identification for the null constant, and it provides a link into the sentence for the discourse topic, via the operator. Therefore, a more accurate structure of this process is as represented in (144) below:



Rizzi (1992) proposes that the discourse operator is intrinsically 3rd person, hence the restriction of object drop in German to 3rd person. In the structure above, the discourse operator has no intrinsic features of its own, but rather gets those features from the discourse topic. Therefore there is no restriction to 3rd person. How do we account for the restriction in colloquial German that Rizzi refers to? While a full explanation of German object drop is not possible here, I believe there is good reason to think that the restrictions on object drop is due to discourse constraints, not syntactic ones.

As we will see in the next section, discourse constraints on empty categories play an important role in restrictions on reference. While object drop in German is restricted to 3rd person, null subjects in Hebrew are restricted to 1st and 2nd person. We will see that the Hebrew restriction on null subjects is due to a preference for topics, subjects, agents and conversational partners. I propose that the restriction to 3rd person for null objects is due to discourse preferences for non-topics, non-subjects, non-agents and non-conversational partners.

In the next section I will present facts about Hebrew null subjects and a theory of discourse identification from Ariel (1990) and Gutman (1999). We will see that Hebrew has person restrictions on null subjects that are accounted for by discourse principles. I will suggest that German null objects are constrained by similar principles. Thus the null operator in German object drop does not have an inherent 3rd person specification. This

is important because we see in Swahili [-SA] clauses, the null operator is not restricted to 3rd person. In fact, we will see that the null subject in [-SA] clauses is restricted to 1st and 2nd person, exactly as in the Hebrew case. The overall conclusion that I wish to argue for is that we need not resort to a stipulation about the nature of the anaphoric operator in order to account for person restrictions in German.

2.14.3 Ariel (1990) and Gutman (1999)

Ariel (1990) discusses the fact that in Hebrew (in the past and future tenses) null subjects are limited to 1st and 2nd person only. She attributes the Hebrew facts to *discourse* restrictions on when a null subject is possible. She claims that antecedents to null subjects are defined along a scale of accessibility that is determined by various factors. We will restrict our discussion to two of these factors: saliency and unity. Saliency is the relative importance an entity has in the conversation. The more salient an antecedent is in the context, the more accessible it is. Topics (i.e., discourse topics) are very salient and hence high on the accessibility scale. Similarly 1st and 2nd persons are more salient (because they are conversational partners) than 3rd person. Thus examples (145a-b, taken from Gutman, 1999) are grammatical in the absence of any context because they are 1st and 2nd person sentences, respectively. However, (145c) is ill-formed because in the absence of any supporting context, the 3rd person antecedent is not salient enough to identify the null *pro*.

- (145) a. *pro* nixshalti ba-mivxan be-historia Hebrew
 failed-1st-sing. In-the-test in-History
 ‘(I) failed the History test.’
- b. *pro* nixshalta ba-mixvan be-Historia
 failed-2nd-sing. In-the-test in-History

‘(You) failed the History test.’

- c. * *pro* nixshal / nixshela ba-mixvan be-historia
failed-3rd-m-sg / f-sg in-the-test in-history
‘(He/She) failed the History test.’

The Saliency Criterion⁴³ includes several ordered pairs, of which (146) shows the more relevant orderings. Thus topics are more salient than non-topics, subjects are more salient than non-subjects, and agents are more salient than non-agents.

- (146) Topics > non-Topics
Subjects > non-subjects
Agents > non-agents

The second factor in determining accessibility is Unity. Unity refers to the level of syntactic/semantic cohesion that exists between two sentences, e.g., conjoined sentences are less unified than a matrix and embedded clause, adverbs can increase semantic unity, etc. An antecedent that crosses a more

⁴³ The Saliency Criterion includes the following ordered pairs, with the element on the left being the more salient.
1st and 2nd persons > 3rd person
subject > object > others
split antecedents interpreted as forming a group > split antecedents not interpreted as forming a group
matrix antecedents > embedded antecedents
discourse-topics > non discourse-topics
antecedents in a Focus Presupposition construction > Antecedents not in a Focus Presupposition Construction.

unified boundary is more accessible.⁴⁴ For example, (147a) is marginal because the antecedent-*pro* relation crosses a sentence boundary that is not semantically unified. In (147b), with the addition of semantic adverbials, unity is increased and thus accessibility is increased.

- (147) a. ?? Noga_i rak hitxatna im Shimon_j Hebrew
Noga only got-married with Shimon

ve-kvar *pro*_{i+j} hitgrashu
and-already got-divorced-pl.

‘Noga just married Shimon, and (they) already got divorced.’

- b. rak lifney xodesh Noga_i hitxatna im Shimon_j
only before month Noga got-married-f with Shimon

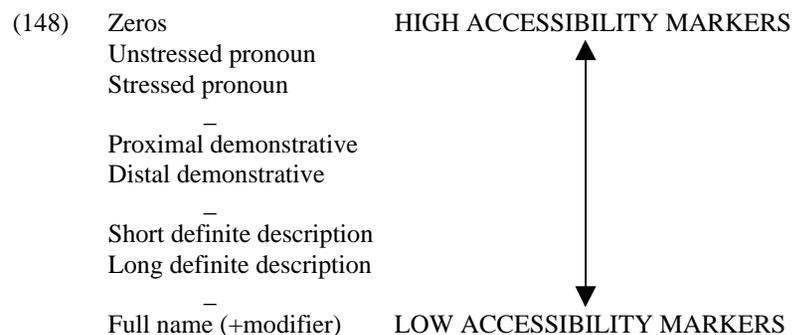
we-kvar ba-shavua she-avar *pro*_{i+j} hitgarshu.
and-already in-the-week that-passed got-divorced-pl

‘Only a month ago Noga married Shimon, and last week (they) already got divorced.’

Ariel also assumes that noun phrases differ in the degree to which they depend on antecedents. For example long definite descriptions occur lower on her scale of accessibility than short definite descriptions, which in turn occur lower on the scale than stressed pronouns, etc. At the highest

⁴⁴ The Unity Criterion includes the following ordered pairs, with the element on the left being the more unified.
Embedding > conjoining
Sentences with parallel time-adverbials > sentences with no parallel time adverbials
Sentences with consequence adverbials > sentences with no consequence adverbials
Sentences with other adverbials > sentences with no other adverbials

end of the scale are gaps, i.e., null subjects and objects. This is shown in (148). Noun phrases at the higher end of the scale will only recover antecedents with a higher level of accessibility (e.g., topics). Similarly, noun phrases at the lower end of the scale can recover antecedents that are lower in accessibility. Therefore gaps, which are the highest in the scale, require the highest degree of accessibility, and hence are the most restricted.



In the case of null subjects, Ariel finds that topics are more accessible than non-topics. This means that null subjects are more likely to occur in contexts in which there is a clear discourse topic available to the listener as an antecedent. In cases where there is no topic, a null subject is not discourse identified. Similarly, subjects are more accessible than non-subjects, and agents are more accessible than non-agents. Therefore null subjects seek out topics, subjects and agents more than other elements to act as antecedents for discourse identification.

How does this theory account for the Hebrew pattern of subject drop? In past and future tenses, Hebrew allows subject drop of 1st and 2nd person pronouns, but not 3rd person pronouns (although see below). Recall

that according to Ariel, null subjects are the highest accessibility markers and thus require an antecedent that is high in accessibility. Ariel claims that 1st and 2nd person antecedents are inherently more salient in the discourse than 3rd person antecedents because they are conversational partners. Because null subjects require the highest level of accessibility, this reduces the frequency of 3rd person null subjects. In fact, Gutman (1999) shows that 3rd person null subjects are not completely unattested, but are considerably less frequent than 1st or 2nd person null subjects. Gutman shows that when a sufficient level of accessibility is created (through increased saliency and unity), Hebrew allows null 3rd person subjects, as shown in (149).

(149) Joan_i soxaxa ita axshav be-ivrit, af ki
 Joan chatted-f with-her now in-Hebrew, even though

*pro*_i hevina rak xelek min ha-dvarim she-ha-yalda amra.
 understood-f only part from the-things that-the-girl said-f

‘Joan_i was chatting with her in Hebrew now, even though (she_i) understood only part of what the girl said.’

Example (149) is taken from a novel, and demonstrates that in an adjunct clause (high in unity) with a matrix subject antecedent (high in salience), pro-drop is possible in the 3rd person. Gutman argues that this is because the antecedent accessibility is extremely high, as well as the fact that this

occurs in literate Hebrew. She argues that the literate medium increases macro (or global) accessibility, making pro-drop much easier.⁴⁵

In addition to this person restriction in past/future tenses, Hebrew disallows null subjects entirely in the present tense. Gutman (1999) argues this is because of an additional condition that impacts null subjects: null subjects must be syntactically identified in order for discourse identification to be possible. Hebrew present tense is unmarked for person features, and so null subjects are not *syntactically* identified. This renders discourse identification irrelevant. In the past/future cases, however, both syntactic as well as discourse identification affect the occurrence of null subjects.

Gutman provides a series of sentences in the present tense with increasing levels of accessibility, and we see that in each case, null subjects are disallowed. In (150), the present tense embedded clause does not allow *pro*, despite a subject antecedent. In (151), the accessibility is increased by incorporating (150) into a conjoined-clause structure with parallel-time adverbials. Because of the parallel time adverbials, an additional clause must be added.

- (150) * Rina_i hodi'a *pro*_i she-magi'a be-shesh
 Rina informed-f that-arrive-f-sg at-six
 'Rina informed that (she) is arriving at six.'
 (lit: Rina informed that (she) is arriving at six.)

⁴⁵ She argues that literate contexts inherently increase saliency, since Ariel's original definition of saliency was based on processing capacity. Ariel argued that the less salient the antecedent the more taxing it is to link to a null argument. In literature, recovery of identity is considerably easier because of the written medium. In fact, even in English in certain literate contexts of extremely high saliency, null subjects are grammatical and very usual. For example, "contains 100% fruit juice" found on a product label. Gutman argues that such examples are licit in English because of the extremely high saliency of the antecedent – the product on which the label occurs.

- (151) */?? etmol Rina_i hodi'a she-hi magi'a be-sheva,
 yesterday Rina informed-f that-she arrive-f-sg at seven
 ve-hayom hi_i hodi'a *pro*_i she-magi'a be-shesh
 and-today she informed-f that-arrive-f-sg at six
 'Yesterday, Rina informed that she would arrive at seven, and today, she informed that (she) would be arriving at six.'

In (152), Gutman adds a preceding sentence that makes the subject into a topic, thus raising the salience even further.

- (152) */? Rina_i hi kol-kax lo hexletit!
 Rina she so NEG decisive
 etmol hi_i hodi'a she-hi_i magi'a
 yesterday she informed-f that-she arrive-f-sg
 be-sheva, ve-hayom hi_i hodi'a *pro*_i she-magi'a be-shesh
 at-seven and-today she informed-f that-arrive-f-sg at six
 'Rina is incapable of making a decision! Yesterday, she informed that she would arrive at seven, and today, she informed that (she) would be arriving at six.'

Example (152) has the highest possible level of salience and unity, and still null subjects are disallowed in the present tense. Thus Gutman concludes that while discourse identification is important in Hebrew, syntactic identification must also be satisfied.

The case of Hebrew highlights the fact that restrictions on person (or number or gender) in null subjects need not necessarily be a result of a syntactic process. In this case we saw that the restriction in Hebrew of null subjects occurring in 1st and 2nd person is due to the saliency of 1st and 2nd person as conversational partners, coupled with the fact that null subjects require a very high level of accessibility.

Let us now consider Swahili [-SA] clauses. Since [-SA] clauses are null subjects, we expect a similar pattern as we see in Hebrew: null subjects have a tendency towards taking 1st and 2nd person antecedents because they are more salient than 3rd person antecedents. In fact, this appears to be the case in Swahili.

Recall that in section 2.11.2 we saw that the reference of [-SA] clauses is free. However, in that section we only discussed the reference of overt subjects. Of the 72 [-SA] clauses in the Swahili corpus, 43 occur with a null subject, and 29 occur with an overt subject. Of the 43 null subject [-SA] clauses, the reference of the null subject was determined on the basis of context. 4 utterances were discarded due to unclear reference. Of the remaining 39 null subject [-SA] clauses, all 39 were either 1st or 2nd person. The overt subject [-SA] clauses, as we saw in section 2.11.2, refer to all three persons. This is shown in table 2.14 below (for examples, refer to section 2.11.2).

Table 2.14 Person reference in [-SA] clauses with Overt and Null subjects

| | Overt Subject | Null Subject |
|------------------------|---------------|--------------|
| 1 st person | 6 | 17 |
| 2 nd person | 7 | 22 |
| 3 rd person | 16 | 0 |

Unclear = 4

We see that restrictions on person occur when the subject is null, but not when the subject is overt. Whether the subject is silent or overt is important in determining any person restrictions in [-SA] clauses. In the absence of an overt operator, the null constant seeks a salient antecedent (in Ariel's terms), hence a preference for 1st or 2nd person (conversational partners). However, if the operator is overt, there are no inherent

restrictions on person. Thus, the restriction to 1st and 2nd person for null subjects in [-SA] clauses is due to discourse principles that apply only to null elements, as described by Gutman (1999) and Ariel (1990) for Hebrew, not an inherent restriction on the operator⁴⁶.

Our conclusion therefore is that [-SA] clauses involve a topic operator – null constant construction. The topic operator can be overt or null, but when null we see the effect of discourse principles on the reference of the null subject.

2.15 Conclusion

In this chapter we discussed the nature of SA, T and OA. We concluded that both SA and OA markers are agreement and not pronominal clitics. We concluded that T is tense and not an auxiliary, although the evidence for this position was somewhat less clear. We then established that Swahili is a null subject language and that in the case of full clauses, the null subject is *pro*, as in Italian and other null-subject languages. We then discussed clauses that permit the omission of SA. We saw that there are two major types of clauses in Swahili that allow SA to be omitted - habitual clauses that do not allow null subjects, and [-SA] clauses that do allow null subjects. The latter raise a problem for the identification requirement on null elements. I proposed that in [-SA] clauses, the subject position contains a null constant licensed by an anaphoric topic operator.

⁴⁶ Intuitions about 3rd person null subject [-SA] clauses vary from speaker to speaker. Of the three native Nairobi Swahili speakers that I have consulted (myself included), the consensus is that 3rd person is "confusing", i.e., there is a strong preference to interpret a [-SA] clauses as 1st or 2nd, and forcing a 3rd person interpretation through rich context conflicts with this preference. This is entirely in-line with the proposal in the text that discourse restrictions prevent 3rd person null subject [-SA] clauses.

This anaphoric operator seeks out a discourse antecedent, to which it links its bindee (the null constant). It is in this way that the null constant is identified. Its reference is not restricted if the operator is overt, but when null, there is a preference for more salient antecedents.

Like other languages, Swahili has null subjects: *pro*, PRO and null constants. Each of these empty categories has distinct properties and is governed by distinct syntactic as well as discourse rules. In chapter 5, I will look at the development of these different types of null elements in child language. We will see that children know the properties of null elements at a surprisingly early age.

The rest of the dissertation is organized as follows. In chapter 3, we will discuss the methodology that I used in collecting, transcribing, organizing and analyzing the child language data. I will describe the subjects, the collection procedures, the transcription protocols, and method of organizing the data into stages, and then the various analyses that I performed. We will then discuss some of the results in chapter 4. I will limit my discussion in chapter 4 to the general pattern of underspecification that is exhibited in the verbal complex by children. I will first discuss several theories of underspecification, and evaluate these theories in light of what we discover about child Swahili. Chapter 5 will then focus on the distribution of subjects in these underspecified clauses.