In chapter 2 we discussed the morphology and syntax of adult Swahili. Particularly relevant to the current chapter is our discussion of null elements in Swahili, and the grammatical omission of SA. We saw that Swahili allows several different null elements: *pro* (the null subject in full clauses), PRO (in infinitival clauses), and NP-trace (in passivization). Since null elements are generally held to an identification requirement, the emergence of clauses that allow SA omission (a salient identifier in full clauses) is intriguing. Adults use habitual clauses which *require* the omission of SA, but as expected by the identification requirement, null subjects are blocked in this environment. Adults also omit SA in declarative, tensed clauses. However, quite unexpectedly, null subjects are allowed in these contexts. We concluded that in these [-SA] clauses the subject position is filled by a null constant that is bound by an anaphoric topic operator. Identification occurs through the anaphoric topic operator, and thus the absence of SA does not block a null subject. Thus subjects in adult [-SA] clauses are in fact the topic operator whose phonetic realization is optional.

After discussing the methodology, data and staging criteria in chapter 3, chapter 4 focused on the omission of inflectional prefixes in child Swahili, and we saw that children omit SA very frequently relative to adults. In the adult grammar, [-SA] clauses occur approximately 5% of the time, while in the child grammar [-SA] clauses occur as frequently as 53% of the time (in stage 2). We also saw that while in the adult grammar no other clause types occur, in child Swahili [-T] clauses and bare stems occur in addition to full clauses and [-SA] clauses. We saw that while these

results are most compatible with ATOM (Schütze & Wexler, 1996; Schütze 1997), there are several questions remaining that ATOM fails to account for.

In this chapter we will investigate the distribution of subjects in the four clause types that occur in early grammar: full clauses, [-SA] clauses, [-T] clauses and bare stems. In particular, we will investigate the use of subjects in the underspecified clauses, looking to see if the theory we postulated in chapter 2 has the expected results. We will see that children adhere to the syntactic restrictions on [-SA] clauses from very early on (i.e., [-SA] clauses only occur in matrix clauses, do not occur with a quantificational operator, optionally take null subjects, etc.). We will also see that expectations of the theory of null constants postulated in chapter 2 is confirmed in the cases of [-T] clauses and bare stems. Our overall conclusion will point to children's ability to discern some very subtle facts in their ambient language about elements that are not even overtly present much of the time. The facts will indicate that they are sensitive to the presence of these elements despite a general poverty of evidence: a conclusion that points to the availability at a very early age of general principles of grammar that regulate the distribution of null elements. The facts will also show that the four clause types are associated with different subject properties, showing that the underspecification of functional prefixes is a syntactic phenomenon.

The chapter is organized as follows. In section 5.1 I discuss several null subject proposals. These are proposals made in the literature to account for why children acquiring non-null-subject languages (such as

English or French) nevertheless produce null subjects. None of these theories were intended to apply to null subject languages, and as such they are not directly applicable to Swahili (a null subject language). However, we might think of SA omission as the Swahili analogue of subject omission. Furthermore, if Zwart (1997) is correct in his proposal that SA is a pronominal subject clitic and not agreement (see chapter 2, section 2.5 for arguments against this proposal), we should see similarities between Swahili SA omission on the one hand, and English null subjects on the other. I first consider a processing theory in section 5.1.1 that says that null subjects in English occur because of children's limited processing capacity. Under this view, the omission of SA in Swahili and the omission of subjects in English are the result of processing limitations. Thus [-SA] clauses should pattern as null subject utterances in English. In section 5.1.2 I consider a PRO theory that claims null subjects in child language are equivalent to null subjects in adult language (i.e., they occur in non-finite contexts). Thus Swahili [-SA] clauses should pattern with other PRO constructions in Swahili. In section 5.1.3 I consider a null-topic proposal that says that null subjects are essentially cases of topic-drop – a process that is possible in many adult languages. I relate this to the proposal made in chapter 2 regarding [-SA] clauses, which involves a topic operator construction. Then in section 5.2 I summarize the relevant facts in adult Swahili (a fuller description of which can be found in chapter 2, section 2) before moving on to subjects in child Swahili in section 5.3. I describe the subject properties of each of the four clause types that Swahili children produce: full clauses (section 5.3.1), [-SA] clauses (section 4.3.2), [-T] clauses (section 5.3.3) and bare stems (section 5.3.4). Section 5.4 then relates these findings to the three null-subject accounts, showing that the processing account cannot be correct. I also show that the null element in subject position cannot be PRO (or *pro*, wh-trace or NP-trace) or a null element resulting from topic-drop. I will show that the proposal in chapter 2 that [-SA] clauses involve a null constant-topic operator construction accounts for the intricate pattern of subject use. Section 5.5 is the concluding section.

5.0   Introduction

Certain languages allow subjects to be null (e.g., Italian, Spanish, etc.), while other languages do not (e.g., English, French, etc.).

(1)   a. Gianni mangia   la mela   Italian
             Gianni eat-3rdsg. the apple
             'Gianni is eating the apple'

          b. Mangia   la mela
             Eat-3rdsg. the apple
             '(He) is eating the apple'

(2)   a. John is eating the apple   English
          b. * is eating the apple

          c. Il mange   la pomme   French
             he eat-3rdsg. the apple
             'He is eating the apple'

          d. * mange   la pomme
                eat-3rdsg. the apple

It has long been noted that young children drop subjects in languages where subjects are obligatory (Hyams, 1986; Hamann, Rizzi & Frauenfelder, 1996; Hamann & Plunkett, 1997; Haegeman, 1995; Rizzi, 1992; Rizzi, 2000).

(3)   a. Ate meat   English (L. Bloom, 1970)
       b. Want more   English (Hyams, 1986)
       c. Oh! Est pour maman – A tout mangé   French
          Oh, is for mummy – has all eaten   (Hamann, *et al.*, 1996)

This observation has led to a considerable amount of research into the omission of subjects in child language. The majority of this research has focused on subject omission in languages that do not allow null subjects because it is in these cases that children diverge from the adult language in a very obvious way. However, less is known about the child's omission of subjects in null subject languages[1]. In this chapter, I will investigate the distribution of null and overt subjects in the early stages of Swahili, a null subject language. In the following section I discuss three proposals in the null subject literature.

5.1   Null Subject Accounts

5.1.1   Processing Limitations

There have been several proposals that claim that processing limitations are the cause of subject omission by young children (L.Bloom, 1970; P.Bloom, 1990; Valian, 1991; Gerken, 199). P.Bloom (1990) argues for a processing solution that makes particular reference to VP length as the determining factor in processing load[2]. His hypothesis is that children have less processing capacity than adults, and hence are forced to omit things that they may well have a full representation for. He argues that the longer an utterance, the greater the processing load. Subject omission

---

[1] An obvious exception is Valian (1991), see the discussion below.
[2] See Hyams & Wexler (1993) for arguments against processing accounts of subject omission in general, as well as specific criticisms of Bloom's (1990) proposal.

should therefore be more frequent in sentences that exert a greater processing load. He predicts that null subject sentences should occur with longer VPs when compared to sentences with overt subjects. He looks at three children taken from the CHILDES database (Adam 2;3-2;7, Eve 1;6-1;10 and Sarah 2;3-2;7) and calculates the length of verb phrases in their utterances with and without subjects.

He calculates VP length by counting the number of words from the verb to the end of the utterance, excluding vocative endings such as *Mommy* or *Daddy*. For example, the sentence in (4) is counted as a VP that is three words long.

(4)   I goed to bathroom, Mommy
              1   2   3

Both mono- and multi-morphemic words are counted as one word, as are simple and compound nouns. Using this measure, Bloom finds that the mean length of VP is significantly lower when subjects are included than when subjects are omitted. For example, for Adam the mean length of VP with past tense verbs is 2.432 (n=44) when the sentences included a subject, and 2.833 (n=36) when the subject was null (Bloom finds that the difference is statistically significant, using a one-tailed t-test). He takes this as evidence that length of VP contributes to the overall likelihood that subjects will be omitted, and thus concludes that processing limitations are the cause of the omission of subjects.

Next, Bloom looks at three kinds of subjects: null, pronominal and full DPs. His hypothesis is that subjects with greater phonetic content create a greater processing load, and so full DP subjects (such as *the boy*) should be more burdensome on the child's processor than pronouns, which in turn should be more burdensome than null subjects. He proposes that full

DP subjects, therefore, should occur with shorter VPs than pronominal subjects, and in turn null subjects[3]. Indeed, this is what he finds for all three children: the VPs in clauses without subjects are longer than those with pronominal subjects, which in turn are longer than VPs in clauses with lexical subjects. Bloom takes this as evidence that the phonetic 'weight' of the subject contributes to processing load, and hence subjects are more likely to be dropped when in sentences that require more processing resources.

Bloom also investigates a claim that subjects are omitted more frequently than objects (Hyams, 1987). Bloom limited his investigation to obligatorily transitive verbs, such as *want* and *pulled*. Below are his results showing the omission of subjects and objects by the three children:

Table 5.1. Omission in obligatory contexts of subjects and objects.

	Adam	Eve	Sarah	Total
Subjects	57%	61%	43%	55%
Objects	8%	7%	15%	9%

As table 5.1 clearly shows, subjects are omitted significantly more often than objects. The question is why should subjects exert a greater processing load than objects? Bloom proposes that the beginning of a sentence poses a greater processing load than the rest of the utterance. Because subjects tend to be utterance-initial, they therefore have a greater tendency to be omitted. However, Hyams & Wexler (1993) point out in their critique of processing accounts that there is ample evidence that the beginnings of words and sentences are perceptually salient for children.

A priori, it seems entirely reasonable to suggest that processing difficulties are at least partially responsible for subject omission. However, as many authors have noted, the processing models proposed by Bloom, as well as Valian (1991), are simply not detailed enough to give a full account of subject and object omission. For example, Bloom claims VP length adds to processing load, but there are several questions about VP length that are left unaddressed: is VP length measured by the number of words, number of morphemes or phonetic weight? Do VPs that contain the same number of words/morphemes but that have different types of words/morphemes (e.g., lexical versus inflectional) exert a different processing load? Do different syllable structures (e.g., CVC versus CVCC) affect processing load differently? None of these questions are addressed by either Bloom or Valian.

5.1.2 Null subjects as PRO

The majority of non-finite clauses (RIs) in several child languages occur with null subjects:

---

[3] It is unclear why *phonetic content* should be the criterion for calculating processing load, especially after the previous calculation of processing load as *number of words in the VP*. After all, as Hyams & Wexler (1993) point out, the standard measure of linguistic complexity has always been morphemes, not phonetic content. Furthermore, if phonetic content were the criterion, we would predict that objects should be omitted as frerquently or more frequently than subjects, as 1st and third person accusative pronouns appear to be as heavy or heavier than nominative pronouns, e.g., Him vs. He → [hɪm] vs. [hi]. However, as Bloom himself and Hyams & Wexler point out, there is a stark asymmetry in the wrong direction between subject omission (roughly 55% at the relevant stage) and object omission (roughly 9% at the same stage) – see table 1 below in the text.

Table 5.2 Null and Overt subjects in non-finite contexts

	Overt	Null	%Null
Dutch (Krämer, 1993)	21	246	92%
German (Behrens, 1993)	278	2199	89%
Flemish(Krämer, 1993)	11	89	89%
English (Phillips, 1995)	34	47	58%

(adapted from Phillips, 1995, tables 14a, 14b, 16, and 18b).

There have been several arguments claiming that early null subjects are instances of PRO (Krämer, 1993; Sano & Hyams 1994), as in the adult examples in (5) below.

(5)   a.   John tried [PRO to climb the stairs]
        b.   PRO to eat fruit is a good thing

In (5a), PRO is in the subject position of the embedded clause and is identified through control from the matrix subject. In (5b) PRO is assigned arbitrary reference.

Sano & Hyams (1994) argue that RIs in languages like French and Dutch are like adult infinitives in the respective adult languages in that they do not raise for agreement and tense feature-checking. They argue that it is precisely in this context that PRO is licensed in the adult grammar. They provide data showing that in English, inflected *be* generally does not occur with null subjects (see table 5.3 below, which can be compared to table 5.4 which shows that overall null subjects are relatively frequent). Sano & Hyams conclude that a fully inflected INFL blocks both PRO as well as RIs, hence accounting for the correlation of null subjects with RIs.

Table 5.3 Null subjects in *is* contexts

	Null subjects with *is*
Adam	11% (13/114)
Nina	4% (2/50)
Eve	0% (0/109)

(Adapted from Sano & Hyams, 1994)

Table 5.4 Overall proportion of Null subjects

Child	Age	Proportion
Adam	2;5-3;0	41%
Nina	1;11.16	44%
	2;2.6	11%
Eve	1;6-2;1	26%

(Adapted from Sano & Hyams, 1994)

A PRO account of null subjects holds great intuitive appeal. After all, the fact that null subjects in child language appear in the same environment as in adult language (in non-finite contexts) points to children having UG principles. However, there are differences in the two contexts that render a PRO account unlikely. First, PRO in child language alternates with overt DPs while in adult language PRO and overt DPs are generally in complementary distribution:

(6)   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]

Second, the reference of child null subjects does not appear to be assigned in the same manner as PRO. It is generally accepted that PRO can receive either an anaphoric reference or arbitrary reference. In the case of control, PRO receives reference from a c-commanding antecedent in an argument position of a higher clause. However, child null subjects do not have arbitrary reference, nor are they controlled by a c-commanding antecedent. In fact, child null subjects occur in root clauses and need not even have a discourse antecedent.

A third argument against a PRO analysis of child null subjects is that the correlation between null subjects and RIs is not as strict as first assumed. It has been documented in several languages that null subjects, while less frequent in finite contexts than non-finite contexts, still do occur

at significant levels. Comparing the rates of null subjects in table 5.2 above and those presented below, it is clear that null subjects occur more frequently in non-finite contexts (table 5.2 above). However, as the numbers below show, null subjects do occur in finite contexts.

Table 5.5 Null and Overt subjects in finite contexts

	Overt	Null	%Null
Dutch (Krämer, 1993)	431	165	28%
German (Behrens, 1993)	2918	781	21%
Flemish(Krämer, 1993)	69	23	25%
English (Phillips, 1995)	79	34	30%

(adapted from Phillips, 1995, tables 14a, 14b, 16, and 18b)

The PRO analysis was the first attempt at understanding the correlation between finiteness and null subjects in child language. While it is appealing in that it assimilates child null subjects to adult language, it faces significant problems as outlined above. We will return to this proposal later in this chapter in evaluating the Swahili data. In the next section I will consider a third proposal that attempts to assimilate child null subjects to an adult process: topic drop.

5.1.3   Topic-drop

There have been several recent topic-drop proposals (Roeper & Rohrbacher, 1994; Bromberg & Wexler, 1995; de Haan & Tuijnman, 1988; Hyams & Wexler, 1993). Bromberg & Wexler (1995) propose a topic-drop analysis on the basis of English data in which they focus on the difference

between null subjects in declarative and wh- contexts[4].

(7) was a green one   decl. null subject   (Eve, 1;10, Brown, 1973)
(8) where do?   Wh- null subject   (Adam, Brown 1973)

Following Sano & Hyams (1994), they propose that there are two mechanisms that give rise to null subjects: a) a grammatical option for children to omit subjects in non-finite contexts similar to adult PRO (cf. Section 5.1.2), and b) an overextension of the adult option to omit topics in certain finite contexts, as initially proposed by de Haan & Tuijnman (1988) for Dutch.

De Haan & Tuijnman propose that subject omission in Dutch is actually a result of the child overextending the contexts in which topics may be omitted (topic drop being a grammatical option in adult Dutch).

(9)   a.   Harry heeft die film   al   gezien   Adult Dutch
                Harry has that picture already seen

       b.   Heeft die film   al   gezien   Adult Dutch
                has   that picture already seen

       c.   [[$_{NP}$ e$_i$]] heeft [$_{NP}$ e$_i$] die film   al   gezien
                    e$_i$ has   e$_i$ that film   already seen

The sentences in (9) are examples taken from de Haan & Tuijnman of adult topic drop. (9a) shows a topicalized subject *Harry*. The standard analysis of SVO subjects in Dutch is that the verb raises to C via T and the subject raises into the [spec, CP] position. Therefore subjects in first position are topics. (9b) shows that it is grammatical for this topic to be dropped, and

---

[4] Bromberg & Wexler essentially replicate an earlier study done by Roeper & Rohrbacher (1994) who were the first to report the wh-OI results discussed below in the text. Nevertheless, I will discuss Bromberg & Wexler's proposal because they are more explicit with respect to the application of their theory to other theories.

(9c) is the structure that de Haan & Tuijnman assign to (9b). Bromberg & Wexler suggest that children allow topic drop in finite clauses in addition to PRO subjects in infinitival clauses.

Bromberg & Wexler discuss Rizzi's (1994) truncation theory (see chapter 4, section 4.2.2). Recall that truncation says that the child has an option to truncate at any projection below CP, thereby accounting for the omission of material in the higher portion of the tree. However, if material high up in the structure is projected, no intervening material may be omitted. Truncation predicts that in the case of wh- questions, root infinitives should not occur since the structure has projected up to the CP layer (as evidenced by the fronted wh- phrase), and hence all intervening projections must be present (crucially including tense). Furthermore, because wh- questions entail a full CP projection, null subjects should also be impossible in wh- contexts because in this case the null subject is not in the specifier of the root (recall our discussion of Truncation and the 'privilege of the root' from chapter 3).

Bromberg & Wexler (1995) investigate the use of null subjects in finite and non-finite wh- contexts to see if this prediction holds. Because topic-drop requires a DP to move into topic position, if a wh-word precedes it then it no longer occupies the specifier position of the root and does not enjoy the privilege of the root. Thus null subjects should not be possible in wh- contexts.

Bromberg & Wexler present data from four English speaking children from the CHILDES database (Adam, 2;3-3;0, Eve, 1;5-2;2, Sarah, 2;3-4;2, Peter, 1;11-2;8). They report two major findings. The first relates to an early stage of development in which null subjects occur frequently in wh- contexts. They show that at early ages, the majority of non-lexical

subjects in both declaratives as well as wh- questions are *null*:

Table 5.6 Proportion of non-lexical null subjects in declarative and wh- utterances at early stages from Adam and Eve (CHILDES, MacWhinney, 2000)

Child	File #	Declaratives	Wh-questions
Adam	11 (2;7.26)	91% (189/208)	94% (15/16)
Eve	1-10 (1;5-1;9)	70% (272/386)	82% (14/17)

(Adapted from Bromberg & Wexler (1995) tables 2 and 4)

Since by their hypothesis topics cannot occur in wh- contexts, subject omission must be due to mechanisms other than topic drop (contra de Haan & Tuijman). The likely candidate is PRO. This predicts that finite wh- questions will occur rarely with null subjects. Indeed Bromberg & Wexler (following Roeper & Rohrbacher, 1994) find that wh- questions with null subjects almost always occur with non-finite verbs. Considering Adam's data (table 5.7, Eve's data is presented in table 5.8), we see that of all the 249 non-finite wh- questions, 118 occur with a null subject (that is 47%), while 2/119 (2%) finite wh- questions occur with null subjects.

Table 5.7 Null and pronominal subjects in finite/non-finite wh- questions for Adam

	Finite	Non-finite
Null Subject	2% (2)	47% (118)
Pronoun Subject	98% (117)	53% (131)
	119	249

Table 5.8 Null and pronominal subjects in finite/non-finite wh- questions for Eve

	Finite	Non-finite
Null Subject	2% (1)	19% (18)
Pronoun Subject	98% (43)	77% (59)
	44	77

In other words, when a wh- question contains a non-finite verb, a null subject is allowed (118/249). However, when a wh- question contains a finite verb, a null subject is not allowed (2/119). This correlation between non-finite verbs and null subjects suggests that in wh-contexts the topic drop option is precluded and only the PRO option is available.

Recall that Rizzi's truncation hypothesis asserts that root infinitives are due to truncation. Under truncation the higher portion of the tree is omitted, but in this case we see two pieces of contradicting evidence: first, non-finite verbs occur in wh- contexts, and second, null subjects occur with the non-finite verbs. This suggests that the omission of at least some subjects is related to the underspecification of tense, and not due to general truncation of structure.

Bromberg & Wexler's second major finding is that at later stages in development, null subjects are considerably less frequent overall, but that when they do occur, they occur in declaratives more frequently than in wh-contexts:

Table 5.9 Proportion of null subjects in declaratives and wh-utterances at later stages from Adam and Eve

Child	File #	Declaratives	Wh-questions
Adam	19 (2;11.21)	16% (28/180)	2% (1/54)
Eve	11-20 (1;10.4-2;2.21)	12% (131/1098)	4% (5/112)

(Adapted from Bromberg & Wexler (1995) tables 2 and 4)

Declarative contexts allow topic-drop while wh- contexts do not. Therefore, at later stages, the preponderance of null subjects in declarative contexts suggests that topic-drop is available, but the absence of null subjects in wh- contexts suggests that the PRO option is no longer available. This suggests that there are two different mechanisms operating for subject omission: topic-drop and PRO.

Summarizing, Bromberg & Wexler conclude that there are two kinds of null subjects in child English. The first involves an adult-like process of topic drop. The second kind of null subject is also adult-like in that it occurs in the absence of finite inflection (and hence can be assimilated to PRO). In order to differentiate these two kinds of null subjects, they look at an environment that disallows topics: wh- questions. In this topic-free environment, we find null subjects occuring in non-finite contexts. This supports the claim that null subjects are licensed in non-finite clauses only, while the occurrence of finiteness forces a topic. It also argues against Rizzi's Truncation Hypothesis because according to truncation not only should there be no Root Infinitives in wh- contexts, we should also not find null subjects in wh- contexts[5].

There are several objections to both the theory and the data that Bromberg & Wexler present. For example, their theory assumes a unitary CP projection, contrary to what is now standardly assumed in the field. Rizzi (1997) gives strong evidence that there are multiple positions above the highest inflectional position (IP, AgrSP or the highest specifier of TP) that correspond to topic positions, a focus position, finiteness position and force position:

(10)   ForceP>TopP*>FocP>TopP*>FinP>IP
         (Where * indicates recursiveness)

This articulated left periphery poses a problem for Bromberg & Wexler because it is crucial to their analysis that there be precisely one left periphery position which when filled by a wh- element is not available for

---

[5] They also note that this fact argues against Bloom's (1990) processing account, which claims that elements at the beginning of sentences are harder for children to process (cf. section 5.1.1).

topic preposing. However, we see in the structure in (10) that there are multiple positions in the left periphery, so that even though wh- elements occur in the focus projection (as noted by Bromberg & Wexler, fn.3, p.227) there are several additional topic positions independently available.

Bromberg & Wexler's data are in contrast to results reported by Valian (1991) who looks at the rates of null subjects in English and Italian children. Valian calculates the number of wh- questions in her corpus of 21 English speaking children. Of the 552 non-subject wh- questions (e.g., *what did Mommy cook?*), she reports that 'the children supplied a subject in all but 9 cases' (p.39). This is a 99% rate of subject use in wh- contexts - in stark contrast to Bromberg & Wexler's numbers. Recall from table 5.9 above that Bromberg & Wexler divided their data into an early stage and a late stage. In the early stage they found frequent null subjects in wh-contexts. However, in the late stage (as shown in table 5.9 above), subjects were almost always overt. Therefore Valian's result may be due to the ages of the children involved.

5.1.4 Valian (1991)

Valian was primarily interested in showing the strength of a processing account over a grammatical account of null subjects of the sort proposed in Hyams (1986). I will not discuss the details of her proposal (which is similar to that of Bloom (1990) discussed earlier), but will instead focus on some of the relevant empirical findings, showing how they are relevant to the topic-drop hypothesis. Her participants were 21 English speaking American children and 5 Italian speaking children (see tables 5.10 and 5.11). She grouped the American children according to MLU (Mean Length of Utterance), as well as verbs per utterance, with the 21 children

falling into 4 developmental groups. She looked at the rate of overt subjects in all non-imitative, non-imperative, usable utterances, and found that in child English overt subjects occur at a very high rate as compared to Italian children. She found that in her least mature group (Group I in table 5.10), subjects occur at a rate of approximately 69%[6]. In the second group subjects occur 89% of the time, and in groups three and four subjects occur 93% and 95% of the time, respectively.

Table 5.10 MLU, age, proportion of verbs and proportion of overt subjects for Valian's American children

	Number of children	Mean MLU	Mean Age	Mean proportion of verbs	Proportion of overt subjects
Group I	5	1.77	2;0	.27	69%
Group II	5	2.49	2;5	.52	89%
Group III	8	3.39	2;5	.70	93%
Group IV	3	4.22	2;7	.79	95%

Table 5.10 shows that American children use overt subjects more than 69% of the time from very early ages. They also exhibit the expected developmental convergence on the adult norm, i.e., overt subjects in nearly all contexts. In fact, Valian examined subjects in parental speech and found that adults consistently use overt subjects between 96% and 98% of the time.

The English results are in marked contrast to the Italian data. The data from the Italian children was divided into two time periods (each time period containing data from all five children). Valian found that overt subjects occur only 20% of the time in Time I and 23% of the time in Time II.

---

[6] As Carson Schütze points out, this rate of subjects is already higher than Adam and Eve's early files, showing that Valian's participants were more mature than Adam and Eve for the files Bromberg & Wexler analyzed.

Table 5.11 Proportion of verbal utterances and overt subjects for Valian's Italian children

	Mean proportion of verbs	Proportion of overt subjects
Time I	.27	20%
Time II	.39	23%

*MLU was not included because of the difficulties in comparing across languages. The children were aged 1;6 or 1;7 at the beginning of the recording period.

While Valian doesn't give the corresponding rates for parents, she cites Bates (1976) who calculated that Italian parents use subjects 30%-40% of the time (Valian's own recounts of Bates' data yield a higher rate of 46%-56%). So while subjects in American and Italian child language do not occur at exactly adult-like proportions, the crucial fact is that Italian children allow null subjects at approximately three times the rate of American children.

The finding that American children allow null subjects at a different rate than Italian children is important because it suggests that the mechanism underlying subject omission in child English is different from that underlying child Italian. Furthermore, since Italian children allow null subjects at a higher rate than Italian adults (according to Valian), it seems that null subjects in child Italian may involve a null subject option as well as a topic-drop option, as Bromberg & Wexler suggest. They propose that null subjects occur because of two independent mechanisms: PRO licensed in the environment of RIs, and topic drop. Since Italian children generally do not produce Root Infinitives (Guasti, 1993/1994; Hoekstra & Hyams, 1998; Rizzi, 1994), this additional discrepancy between adult and child Italian subjects may be attributed to topic drop. So while Valian found very few null subjects in wh- contexts in English, in contradiction to Bromberg & Wexler's results, her Italian findings are compatible with their analysis.

5.1.5   Wang *et al*. (1992)

Along the same lines, Wang, Lillo-Martin, Best & Levitt (1992) investigate null subjects and null objects in Chinese and American children. They address a proposal by Jaeggli & Hyams (1988) in which English children are hypothesized to have mis-set the null subject parameter. Jaeggli & Hyams propose that English children initially assume that their language allows null subjects that are identified through discourse (as in Chinese). They later acquire the English setting and null subjects cease to be a possibility at this point.

Adult Chinese allows both null subjects (approximately 36% of the time) and null objects (approximately 10% of the time), while English disallows both. Wang *et. al*. report that both American and Chinese 2-year olds allow null subjects, but Chinese children drop subjects 46% of the time, while American children omit subjects 14% of the time[7]. This difference is statistically significant.

Furthermore, the difference between Chinese children and adults is not significant. They argue that the difference between Chinese children and American children on the one hand, and the similarity between Chinese children and Chinese adults on the other, shows that the mechanism underlying early English null subjects is distinct from that of Chinese.

Wang *et. al.* also report that Chinese children omit objects approximately 22% of the time, while English speaking children drop

---

[7] The overall proportion of null subjects for American children is 33%, but some of these occur in embedded infinitival contexts and are thus grammatical. They adjust this by eliminating the contexts in which the null subject is grammatical in English and find that 14% of obligatory subjects are null. For the Chinese children, since Chinese is a null subject language, all null subjects are grammatical.

objects only 3.75% of the time (see table 5.12 below). This difference is statistically significant. They argue that if English speaking children are learning a Chinese-type language, then we expect Chinese children and American children to allow null objects at roughly similar frequencies. Because this is not the case, they conclude that the mechanism underlying null objects in child English is distinct from that underlying null objects in Chinese.
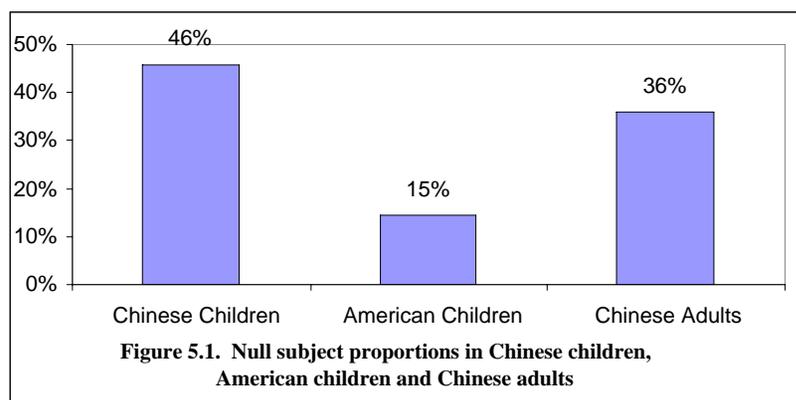

Figure 5.1. Null subject proportions in Chinese children, American children and Chinese adults

Table 5.12 Chinese and American Children's proportions of null arguments

	Chinese Children	American Children
Mean % of null Subjects	46.5%	14.6%
Mean % of null Objects	22.5%	3.75%

Adapted from Wang *et.al.*'s figures 1 and 2

Thus Wang *et.al.* argue on the basis of frequency differences that the mechanism for null arguments in child English is different from that of either child or adult Chinese. This is reminiscent of Valian's (1991) methodology, where she argued that child English null subjects are distinct from Italian null subjects on the basis of frequency differences. A criticism

of this approach is that frequency differences are difficult to interpret. For example, the null subject differences that are observed in early Chinese versus early English may be attributed to discourse differences in the two languages. It may be that discourse salience in the two languages differs because of different morpho-syntactic properties of the two languages, leading to different proportions of discourse-identified null subjects.

Wang *et. al.* reject a purely discourse account on the basis of the subject-object asymmetry. The relative frequencies of null arguments show that while both null subjects and objects are permitted in early Chinese, only null subjects are permitted in early English (Wang *et.al.* take the 3.75% of null objects in American children's speech to be speech errors). They propose that there are two parameters that account for these differences: the Discourse-Oriented Parameter which permits discourse-oriented languages (in the sense of Huang, 1986) to have null arguments, and the Null Pronoun Parameter, which allows null arguments licensed by case-assigning categories and identified by rich agreement (Rizzi, 1986). The basis of this dichotomy is the relative frequencies of null arguments. While English children use significantly fewer null subjects than Chinese children (14.6% versus 46.5%), they nevertheless do use null subjects. However, as we see in table 5.12, null objects are absent in the child English. They claim that because null objects are virtually unattested in English, this points to a grammatical restriction.

So far we have discussed three accounts of child null subjects. The first approaches the problem from a processing perspective, claiming that young children suffer from processing limits which force them to omit certain elements, in this case, subjects. The second approach takes

advantage of the fact that null subjects in non-null subject languages tend to occur in Root Infinitives. This draws a close parallel between child RIs and adult infinitives in that the lack of finiteness provides a licensing context for null subjects. Therefore, a natural conclusion is that child null subjects are licensed as they are in adult infinitives, i.e., they are PRO. The third proposal claims that there are two mechanisms that underlie the omission of subjects: a grammatical option to use null subjects related to the RI phenomenon (i.e., PRO), and an adult-like option of a null topic in [spec, CP]. We then discussed issues of frequency, noting that relative frequency of null arguments may be suggestive of different grammatical mechanisms.

In the remainder of this chapter we will investigate the use of subjects in child and adult Swahili. While Swahili does not have wh-fronting (and hence we cannot test the topic-drop hypothesis in the manner that Bromberg & Wexler did), I will provide evidence that supports a topic-drop analysis of certain Swahili null subjects. The analysis I propose, however, is considerably different from Bromberg & Wexler's. I make use of the syntactic analysis discussed in chapter 2 in which the omission of SA involves a topic operator that binds a null constant in subject position. I will show that the crucial features of this construction (as outlined in section 2.14) hold for child language as well. The conclusion we reach is that children are remarkably sensitive to the syntax of their language and to general principles of grammar, specifically the restrictions on null elements (such as null constants and null operators). Not only do they acquire these elements early in the acquisition process, they adhere to the restrictions that they impose to a surprising degree.

The remainder of this chapter is organized as follows. In section 5.2, I will discuss the distribution of overt subjects, null subjects, and topics

in the adult grammar of Swahili in the four clause types described in chapter 3 (Full clauses, [-SA] clauses, [-T] clauses and bare stems). This will establish the baseline for this dialect of Swahili. I will argue that the different clause types have different subject/topic properties, specifically that null subjects in full clauses are *pro* and in [-SA] clauses are null constants bound by a null topic. I will show that [-T] clauses and bare stems are exceptionally rare in adult speech, and hence cannot be compared to child language. In section 5.3, I will investigate overt subjects, null subjects, and topics in child language. I will show that there are differences between the various clause types with respect to how subjects distribute and that the same syntactic principles which operate in the adult grammar operate in the child grammar as well. I will examine the rates of null subjects in the various clause types and compare them to adult rates (as Valian, 1991 did). In section 5.4 we return to these theories of null subjects and evaluate them with respect to the Swahili data. I conclude in section 5.5.

5.2   **Adult Swahili**

In this section I will discuss overt and null subjects in various clause types in adult Swahili. I will first show that full clauses in adult Swahili allow null subjects freely. I argued in chapter 2 that these null subjects are *pro*, as in other well-known null subject languages. I will then discuss the occurrence of [-SA] clauses and discuss some of the discourse restrictions on these clauses. I argued in chapter 2 that subjects in [-SA] clauses are topics that bind a null constant in subject position. I will then discuss the occurrence of other underspecified clauses in adult Swahili,

concluding that they are exceptionally rare. In section 5.3 we turn our attention to subjects in child Swahili.

5.2.1   Swahili is a null subject language

As we saw in Chapter 2, section 2.9, Swahili is a null subject language. Traditionally, rich subject agreement (which marks person and number) has been thought to license null subjects, since the identity of the missing subject can be recovered from this verbal morphology (Taraldsen, 1978; Rizzi, 1986).   Swahili agreement is rich in that it marks person and number (see chapter 2, section 2.4.1). I argued in chapter 2 that the null subject in Swahili full clauses is *pro* which is licensed by subject agreement, while the null subject in Swahili [-SA] clauses is a *null constant*, bound by an anaphoric topic operator, in the sense of Rizzi (1992) and Rizzi (1997). We will return to this point shortly.

To my knowledge there are no corpus based studies which have documented overt-to-null subject rates in adult Swahili. In order to establish this rate for Swahili, I coded parental utterances in 16 files (4 of Haw's files, 8 of Mus' files, and 3 of Fau/Has' files). This allowed me to establish a quantitative baseline for the 'target' language that the children were exposed to. Using the COMBO tool in CLAN (MacWhinney, 2000), I counted the frequency of subjects in adult speech in this corpus. I counted as subjects: lexical NPs, pronouns, names and demonstratives, all in preverbal position. In the case of demonstratives, each item was examined individually and a determination made whether or not the demonstrative was in fact acting as a subject, an adjunct or a topic. An example of each is given below in (11)-(13).

(11)   hii imeribika   Subject
        this 3rd inanim.- pr.prf..- spoil – IND
        'This has spoiled'

(12)   huko alianguka   Locative Adjunct
        There 3rds – past – fall – IND
        'He fell there'

(13)   hile ninafikiri imeribika   Topic
        That 1sts–pres–think–IND 3rdinanim.–pr.prf.–spoil–IND
        'As for that thing, I think it is spoiled'

Only examples such as (11) were counted as subjects, where the demonstrative *hii* 'this' is the logical subject, and agrees with the verb in person/number/noun class.   In the last two cases, the NPs are in preverbal position, but are not counted as subjects.   Example (12) shows an adjunct locative demonstrative fronted for focus (the unmarked order is *alianguka huko*), while example (13) shows a topicalized demonstrative, as indicated by the English translation.

Furthermore, whenever more than one subject-like element (nouns, pronouns, names, demonstratives) occurred in preverbal position, a determination was made whether one or none were subjects. The criteria for this determination were context, matching with the agreement morphology, and intonation. If so-called 'comma intonation' was used (an intonational break between the preverbal DP and the rest of the sentence), the preverbal element was classified as a topic. The example below shows an interaction between an adult (Ali) and a child (Fau), in which Fau uses the demonstrative to specify an item (ball) in her question to Ali. In his response, Ali uses the demonstrative to refer to the ball, and he focuses 'the ball' by moving it to the front of the sentence.

(14)   Fau: nani alichukua hile ball?
                Who 3rds – past – take – IND that ball
                'Who took that ball?'

        Ali: hile, Hassan alichukua.
                That Hassan 3rds – past – take – IND
                '(as for) that, Hassan took it'

In this case, *hile* carries a lower intonation and there is a slight pause before *Hassan*. This is typical 'comma intonation'. *Hile* and *Hassan* can not be a single constituent (i.e., a demonstrative used to pick out Hassan) because of this intonation, and using the context from Fau's utterance, it is clear that *hile* refers to the ball. Thus, in this case, this utterance was counted as containing a topic (*hile*) and an overt subject which was the name *Hassan*. The results of this analysis are presented in table 5.13:

Table 5.13 Overt and Null subjects in the speech of adults

	Overt	Null	
Adult subject use	263 (17.8%)	1207 (82%)	1470

The results show that the three adults in this corpus use overt subjects 17.8% of the time overall. Therefore null subjects predominate. As in other *pro*-drop languages, overt subjects are usually used to disambiguate referents, to focus the subject, to contradict previous utterances, etc. For example:

(15)   Speaker 1:   Juma   a – li–shind – a?
                      Juma SA-past-win-IND
                      'Did Juma win?'

        Speaker 2:   Hapana, Mariam   a – li–shind – a?
                        No   Mariam SA-past-win-IND
                      'No, Mariam won'

                      *Hapana,   a – li–shind – a?
                        No   SA-past-win-IND

In example (15), speaker 1 asks a question, and in speaker 2's negative response, the subject must be specified. Without the subject the sentence is ungrammatical on the intended interpretation. Conversely, if speaker 2 responds in the affirmative, it is pragmatically odd for the subject to be specified:

(16)   Speaker 1:   Juma   a – li–shind – a
                      Juma SA-past-win-IND
                      'Did Juma win?'

        Speaker 2:   ??Ndio,   Juma   a – li–shind – a
                        Yes   Juma SA-past-win-IND
                      'Yes, Juma won'

                      Ndio,   a – li–shind – a
                      Yes   SA-past-win-IND
                      'Yes, (he) won'

Instead, the more natural response would include a null subject. Similarly, when disambiguating subjects, a subject is required[8]:

(17)   Speaker 1:   Nani   a – li–shind – a, Juma au Mariam?
                      Who SA-past-win-IND, J. or M.
                      'Who won, Juma or Mariam?'

        Speaker 2:   Mariam   a – li–shind – a?
                      Mariam SA-past-win-IND
                      'Mariam won'

                      *a – li–shind – a?
                       SA-past-win-IND
                      '(He/she) won'

In the above example, because Swahili SA does not distinguish between masculine and feminine, an overt subject is required to disambiguate the referent. A null subject sentence is pragmatically infelicitous. Furthermore, overt subjects occur contrastively, as in the following example:

---

[8] The name *Juma* is a male name and *Mariam* is a female name.

(18)   Speaker 1:   Juma   a – li–shind – a
                         Juma SA-past-win-IND
                         'Juma won'

         Speaker 2:   Hapana, Mariam   a – li–shind – a
                           No   Mariam SA-past-win-IND
                         'No, Mariam won'

                         *Hapana,   a – li–shind – a
                            No   SA-past-win-IND
                         'No, (he/she) won'

In this example, a null subject is infelicitous. All the examples given so far involve the alternation between null subjects and proper names. However, typically, pronouns alternate with null subjects. Pronouns alternate with null subjects based on contrast, or as Ashton (1947) puts it: "(pronouns) may be used with either subject or object prefix to give emphasis" (p.44).

(19) Mimi   ni – me – kwish – a , lakini yeye h – a – ja – kwish – a
       I SA_{1s}–p.perf.–finish–IND but   he NEG–SA_{3s}–neg-perf.–finish–IND
       'I have finished, but he hasn't finished (yet).'

(20)   ? ni – me – kwish – a , lakini h – a – ja – kwish – a
         SA_{1s}–p.perf.–finish–IND but NEG–SA_{3s}–neg-perf.–finish–IND
         'I have finished, but he hasn't finished (yet).'

Example (19) above shows a case of contrastive use of a pronoun. The corresponding null subject version in (20) is awkward because it assumes that the contrastive nature of the two referents is clear in discourse. This of course is possible, but dispreferred in the absence of clear supporting discourse.

It is in such contexts in which adult Swahili speakers use overt subjects. Such cases aside, when the subject of the sentence is clear from discourse, then it is preferred to omit the subject, as in example (16) above.

5.2.2   [-SA] clauses for adults

Recall from chapter 4, section 4.4.4, children produce [-SA] clauses. These are verbal utterances which are missing subject agreement, but have tense (and optionally object agreement). This is schematized again in (21) (ignoring OA) :

(21)   [-SA] clause: Ø – Tense – Verb – Final Vowel

We discussed in chapter 2 that in Standard adult Swahili such sentences are ungrammatical, but in Nairobi Swahili, in very restricted contexts, adult speakers drop SA marking. Just as in child [-SA] clauses, these clauses have no SA prefix but are fully marked for tense and optionally other affixes. Examples (22)-(24) are adult utterances directed to children, taken from the Swahili corpus.

(22)   na – tak – a ice?   Hami, HAW02
           Ø pres-want-IND ice?
           'Do (you) want some ice?'

(23)   weh   na – ju – a   ku – onge – a?   Ala, MUS09
           You Ø pres-know-IND inf-speak-IND
           'Do you know how to speak?'

(24)   ta – ku – chun – a   Mot, MUS10
           Ø fut–OA_{2s} – pinch–IND
           '(I) will pinch you'

In (22), there is no SA and the verbal complex begins with the present tense marker *na*. The SA that would occur in a full clause with this meaning would be *u*, SA_{2s}. (23) contains an overt subject (*weh*, a reduced form of the 2nd person singular pronoun *wewe*), and SA is missing. In (24) the future tense marker *ta* begins the verbal complex, and there is OA but no SA. The SA that would occur in a full clause with this meaning would be *ni*, SA_{1s}. (25) contains OA, and an applicative suffix, but no SA.

(25) ta – ni – nunu– li – a   nini?   Ali, FAU02
      fut–OA_{1s}–buy–applic–IND what
      'What will (you) buy for me?'

[-SA] clauses of this type occur in adult Nairobi Swahili when the subject is understood through discourse, and are only permissible in spoken language. In chapter 3 I showed that children use [-SA] clauses quite frequently (more than 50% of the time in stage 2, and almost 30% of the time in stage 4). Although such clauses occur in the adult language, their overall frequency is considerably lower. Of all adult verbal utterances in this corpus, 4.9% occur without SA marking (excluding imperatives, repetitions, etc.). Since this is ungrammatical in Standard Kiswahili, it is possible that these [-SA] clauses constitute an adult error. In order to test this, I took digitized segments of speech from the corpus which contained [-SA] clauses and presented them to a native speaker of Standard Kiswahili who was also fluent in the Nairobi dialect of Swahili. I presented him with 10 segments of speech each containing one [-SA] clause uttered by an adult with context. In all cases, the consultant judged the sentences grammatical in informal and colloquial contexts. I thus conclude, based on the judgments of the native speaker consultant as well as my own judgments, that [-SA] verbs are a restricted but grammatical option in adult spoken Nairobi Swahili. In chapter 2 we discussed the syntactic nature of [-SA] clauses and concluded that they involved a null constant bound by a topic operator. In this section we will discuss the context in which [-SA] clauses occur.

In determining the context of the [-SA] clauses, it is somewhat easier to do with adult data than child data because the context is usually a lot richer, and no other omissions occur (neither morphological nor lexical). While a full discourse analysis of these data is beyond the scope of this dissertation, I have some preliminary observations about SA omission in adult Nairobi Swahili.

The omission of SA in adult Swahili is subject to three related principles. First, SA may never be dropped if the topic/subject is not extremely salient. Only in the cases when there is no doubt or ambiguity over the identity of the subject of the sentence can SA be dropped. This may be established by previous discourse, physical gestures, eye gaze, prior shared knowledge, etc. Second, the topic (if there is one) must have the same identity as the subject. That is, direct and indirect object topics do not occur in such contexts[9].

(26) a.   Mimi Ø – ta – nunu – li – a wa-toto chai
             I   fut–buy–applic–IND 2–child tea
             'I will buy some children some tea'

      b.   * Watoto Ø – ta – nunu – li – a   chai
               2-child   fut–buy–applic–IND tea
             'Some children, (I) will buy (them) some tea'

      c.   *chai Ø – ta – nunu – li – a   wa-toto
             tea   fut–buy–applic–IND 2-child
             'Some tea, (I) will buy some children'

---

[9] An objection to the examples given in the text is that they are ungrammatical because of the lack of OA. Some have argued that OA is obligatory whenever the object is topicalized. Recall that OA occurs when the object is *specific*. Therefore the requirement that OA be present in topicalized structures may be because topicalized information is usually known information, and hence usually specific. However, it is possible to topicalize known information that is non-specific. The example in the text is ungrammatical even under the reading of the object being non-specific. Furthermore, even if we add OA, the examples are ungrammatical:
*watoto Ø – ta – wa – nunu – li – a   chai
 2-child   fut–OA_2–buy–applic–IND tea
 'The children, (I) will buy (them) some tea'

Third, [-SA] clauses can only be used within episodes. An episode is defined as 'the set of sentences with the same (null) topic'. Therefore [-SA] clauses can never be used to mark an episodic boundary.

(27) A1:   Rafiki yako, a – na – it – w – a – je?
            Friend yours SA_{3s} – pres – call – pass. – IND – Q
            'Your friend, what is he called?'

      B1:   Juma

      A2:   Ø – na – fany – a kazi wapi?
              pres – do – IND work where
            'Where does (he) work?'

      B2:   Huko Dandora
            There Dandora

      A3:   Ø – na – ish – i huko Dandora?
              Pres – live – IND there Dandora
            'Does (he) live in Dandora?'

      B3.   Ehh
            Yes

      A4:   * Na wewe, Ø – na – ish – i huko pia?
             And you,   pres – live – IND there also
             'And you, do (you) live there too?'

The example in (27) is a dialogue between two individuals labeled A and B. Each utterance is labeled by speaker as well as numbered sequentially for ease of reference. A1 starts the dialogue by setting the topic as *rafiki yako*, 'your friend'. B1 answers the question by giving the friend's name, *Juma*. A2 continues within the episode, questioning where Juma works. This utterance is grammatical without SA because it is within the episode (i.e., the topic of discussion is still *Juma*), and the subject is co-referent with the topic. B2 answers the question with simple location *Huko Dandora*. A3

continues the episode with a question about whether *Juma* lives in *Dandora*, and again SA omission is grammatical. However, A4 signals a shift in episodes because the topic moves from Juma to one of the interlocutors (B). Since A4 is an episodic boundary, SA omission is ungrammatical.

It must be stressed that these are preliminary observations about when SA may be omitted in adult Swahili. Nothing in the way of a concrete theory is being presented because the adult data are too sparse and the judgments that I am relying on are primarily my own. My consultant found producing these clause types in artificially constructed contexts difficult. This remains an important area of future work.

5.2.3   **Other clause types**

The fact that [-SA] clauses occur approximately 5% of the time suggests that perhaps other clause types that children produce also occur in adult speech. However, [-T] clauses and bare stems combined constitute only 1.2% of adult verbal utterances. These sentences were clearly speech errors or ellipsis. Speech errors were determined by the context, e.g., if tense was omitted in a subjunctive context, it is likely that the intended utterance was a subjunctive verb (recall that subjunctives are unmarked for tense in Swahili), and the adult mispronounced the final vowel. Examples of such utterances are given below:

(28)   tu – end – a   ku – swali?   Ala, MUS13
       SA_{1pl}–go– IND   inf – pray
       'Should we go to pray?'

(29)   tu – Ø – mw–ambi–a   Faiza   a – nunu – e   juice?   Ali, FAU03
       SA_{1pl}–Ø–OA_{3s} – tell–IND   Faiza SA_{3s}–buy–SUBJ juice
       'Should we tell Faiza (that) she should buy juice?'

In (28) the intended meaning is a question/suggestion, and the appropriate adult sentence would have included a subjunctive final vowel on the matrix verb. Instead the adult here uses an indicative final vowel, but also omits the tense marker. Because the context fits the subjunctive interpretation, this is more likely to be a mispronounced final vowel than a tenseless root clause. In (29), notice that in both the matrix as well as the embedded clause the subjunctive is intended, and so in all likelihood this utterance is also a mispronounced final vowel in the matrix clause. An example of a bare verb by ellipsis (the only example of its sort in the corpus) is given in (30):

(30) Adult: u – na – tak – a   cha – kula?
              SA_{2s}–pres–want–IND   7-food
              'Do you want food?'

      Mus: …   (no response)

      Adult: hmm?

      Mus: …   (no response)

      Adult: Mustafa?

      Adult: u – na – tak – a   cha – kula?
              SA_{2s}–pres–want–IND   7-food
              'Do you want food?'

      Adult: tak – a   cha – kula?   (raised voice)
              Want-IND 7-food
              'Want food?'

In this interchange, the adult asks the child a question using a full verbal complex and the child doesn't respond. The father asks again, and the child still doesn't respond, and the father in his third repetition of the question (in

a raised voice) uses a bare stem. My judgment is that the bare verb is ungrammatical, but is acceptable in this context. My other native speaker consultant agrees. This is similar to the hypothetical English case in (31) in which the tense bearing element *do* is omitted:

(31) Adult: do you want to eat?
      Child: …
      Adult: hmm?
      Child: …
      Adult: Hey, do you want to eat?
      Adult: want to eat?
      (raised voice)

The relative proportions in adult speech of the various clause types are given in table 5.14.

Table 5.14 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

Figure 5.2. Frequency of Overt Subjects in adult Full Clauses and [-SA] Clauses
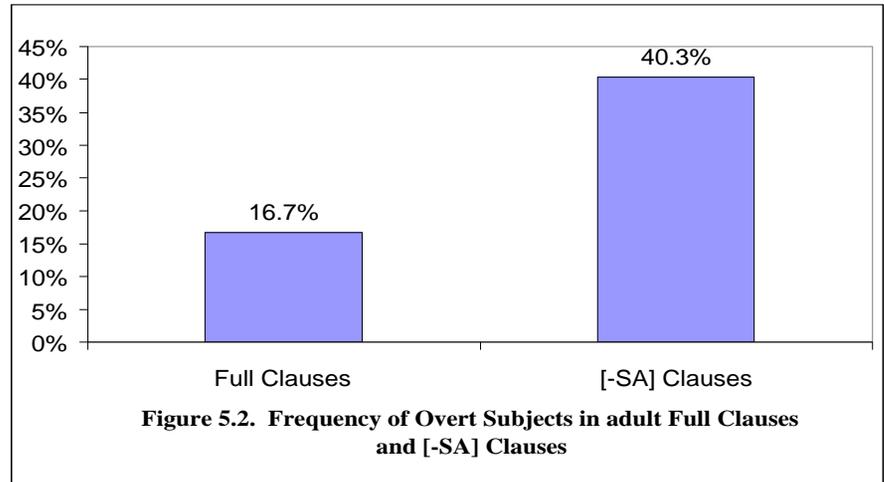
Because of the low frequency of [-T] clauses and bare stems and the judgments from native speakers, I assume that they are not grammatical in

the adult language. On the other hand, as noted earlier, [-SA] clauses are grammatical in the adult language, subject to strict discourse and pragmatic constraints.

As we will now show, the grammaticality of [-SA] clauses is supported by the fact that there are syntactic effects of this underspecification of SA. Subject agreement is related to null subjects: as mentioned earlier, Taraldsen (1978) and Rizzi (1982; 1986) propose that null subjects in *pro*-drop languages are licensed by rich agreement. Therefore the omission of rich SA in Swahili should have an effect on null subjects. Our prediction is that the omission of SA results in the obligatory use of subjects, since null subjects are no longer licensed. Therefore the rate of subject use becomes important not only in adult speech in general, but specifically in [-SA] clauses.

5.2.4   Subjects in different adult clause types

Recall that we found that overall, approximately 18% of verbal utterances contain overt subjects (see Table 5.13). However, we now know that there are at least two types of grammatical clauses in adult Nairobi Swahili: full clauses and [-SA] clauses. An automated CLAN analysis reveals that in full clauses, overt subjects occur approximately 16.7% of the time (230 out of 1380 full clauses had an overt subject). However, in [-SA] clauses, overt subjects occur approximately 40% of the time (29 out of 72 [-SA] clauses had an overt subject). This difference is statistically significant (_-level = 0.01, $\_^2$ = 14.296) according to a $\_^2$ test (with Yates Correction Factor).
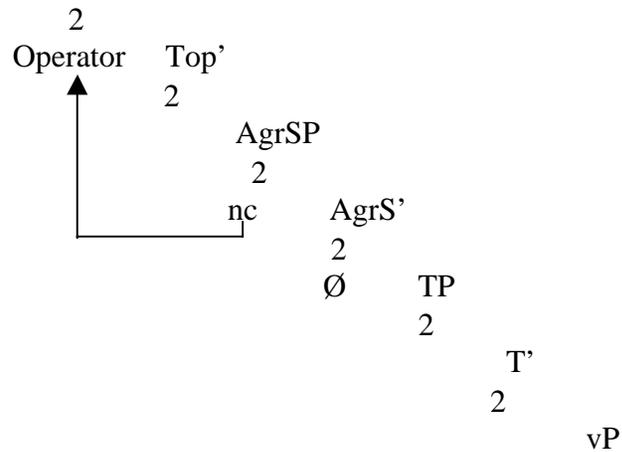
Thus the prediction that [-SA] clauses should only occur with overt subjects is false. Summarizing what we have found so far, overt subjects

occur rather sparingly in adult Swahili: only 18% of all sentences have an overt subject. Furthermore, the proportion of overt subjects is considerably higher in [-SA] clauses than in full clauses.

5.2.5   Topics or Subjects

Recall from chapter 2 section 2.5 that Zwart (1997) and Buell (1999) argue that what look like subjects in Swahili are in fact topics, and that the SA marker is not a realization of agreement, but rather a reduced pronoun in subject position. I presented a different perspective on subjects and topics. I claimed that overt subjects in Swahili full clauses are always true subjects (i.e., occur in [spec, IP], in pre-Minimalist terms) but that in [-SA] clauses, preverbal nouns are not subjects, but topics. Thus, the difference in the rate of "subjects" in full clauses versus [-SA] clauses suggests two different underlying mechanisms: in one case we are dealing with subjects, and in the other topics. The structure that I proposed for each clause type is given below:

(32)   AgrSP
       2
Subject   AgrS'
         2
       SA   TP
         2
            T'
           2
            vP
            2
              v'
              2

a)   Full Clause

```
      2
Operator    Top'
            2
               AgrSP
               2
          nc      AgrS'
                  2
                 Ø      TP
                        2
                           T'
                           2
                              vP
```
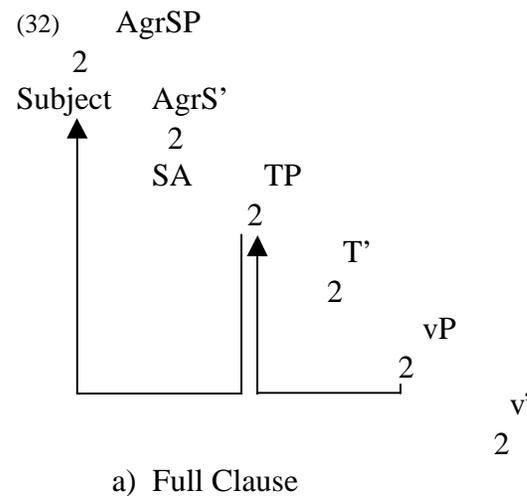
b). [-SA] clause

In full clauses, the subject raises from [spec, VP] through [spec, TP] and into [spec, AgrSP]. In [-SA] clauses, the subject position is filled by a null constant which is bound by an optionally null anaphoric topic operator. The null constant checks the Case feature and satisfies the EPP. According to Rizzi (1997), the null constant requires a topic (for identification) even though the topic may not be pragmatically necessary: topics may be dropped for various discourse/pragmatic reasons, e.g., see Shibamoto (1983) for a description of topic drop in Japanese.

With this understanding of the adult system, we are now in a position to investigate subjects in child Swahili. Specifically, we can investigate the following questions. What is the overall rate of overt subjects? How do subjects develop across time? And how do subjects distribute across clause types? What can the different subjects tell us about the inflectional structure of the different clauses and about children's

knowledge of grammatical principles? In what follows I will investigate each of these questions in turn. For ease of exposition, I will use the term 'subject' to refer to any preverbal DP and not make the distinction between subjects and topics, except when it becomes relevant.

5.3   Subject use by children

In looking at subjects in child Swahili, a useful place to start is to compare the rates of overt and null subjects to the adult rates. Using CLAN, a COMBO analysis of overt subjects in adult and child Swahili found that adults do not differ dramatically from children. Figure 3 presents aggregate data from all the children (n=4) and three adults, showing the proportion of overt subjects in all indicative clauses. As can be seen from figure 3, the overall difference between adults and children is not very large. Children tend to use slightly more overt subjects than adults, but not significantly so.

Table 5.15 Overt and null subjects in child and adult Swahili (pooled across all files)

	Children	Adults
Overt Subjects	269	226
Null Subjects	1013	1090
	1282	1316

Recall that in adult Swahili, subjects[10] occur in different proportions in full clauses and [-SA] clauses (17% and 40%, respectively).

---

[10] Here I use the term subject to refer to a true subject in a full clause, and a topic in a [-SA] clause. See chapter 2, section 2.14 for details on the analysis of [-SA] clauses.
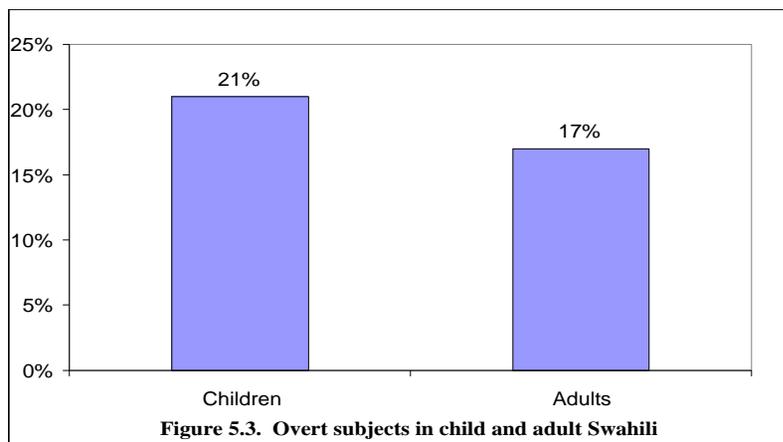
Figure 5.3. Overt subjects in child and adult Swahili

Recall further from chapter 3 that in child Swahili we find four clause types: full clauses, [-SA] clauses, [-T] clauses and bare stems. We will now investigate subjects in different clause types in child Swahili.

In the next four sections I will present results of analyses of subjects in each of the four clause types in child Swahili. I will conclude that lexical subjects and *pro* are permitted in full clauses only, while the null constant – topic operator construction is permitted in [-SA] clauses and bare stems only. I will also show that [-T] clauses do not permit any overt subjects, allowing only PRO subjects. These results are summarized in table 5.16 below.

Table 5.16 Summary of clause types and "subject" options

	Full Clause	[-SA] Clause	[-T] Clause	Bare Stem
Pro	a	*	*	*
Lexical Subject	a	*	*	*
n.c. + Topic	*	a	*	a
PRO	*	*	a	*

5.3.1 Full clauses

Full clauses contain verbs which have all the required prefixes, namely SA and T. The overall rate of overt subjects in full clauses is 23% (119/511). When we look at the rate of subjects in child full clauses across time (figure 5.4), we find that the rate of overt subjects is very close to that of adults in early stages, and then rises after stage 2. This increase in subjects is not significant according to a McNemar's Test (z=0.7921).
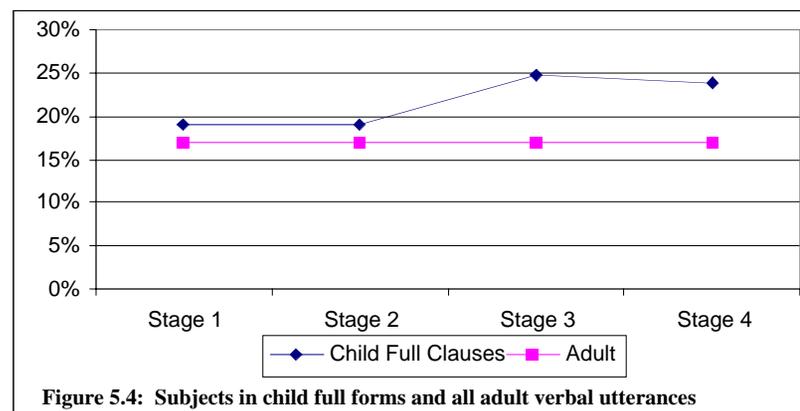

Figure 5.4: Subjects in child full forms and all adult verbal utterances

Examples of children's full clauses with subjects are given below:

(33) pia huyu   a – na – va – a   vi – atu   FAU02 line 988
      also he SA_{3s}–pres–wear–IND   8 – shoes
      'Even he is wearing shoes'

(34) MOT: we   u – na – tak – a   ku–end–a   Ushako ?
            You SA_{2s}–pres–want–IND   inf–go–IND   Ushako
            'Do you want to go to Ushako?'

      HAS: eh, na Sauma … a – ta–end–a   Sago   HAS01 line 194
            Yes and Saumu SA_{3s}–fut–go–IND Sago
            'Yes, and Saumu will go to Sago'

(35)   na wewe   u – na – ruk – a   HAW07 line 1460
         and you SA$_{2s}$–pres–jump–IND
         'And you are jumping (down)'

(36)   MOT: Sumaya   a – li – tow – a ?
                Sumaya SA$_{3s}$–past–remove–IND
                'Did Sumaya remove (it)?'

         (MUS shakes his head)

         MOT: hmm?

         MUS: mimi… ni – li – tow – a   MUS05 line 1543
                 I   SA$_{1s}$–past–remove–IND
                'I…removed (it)'

5.3.2   [-SA] clauses

Moving on to [-SA] clauses, we find that as in the adult grammar, overt subjects occur more frequently in [-SA] clauses than in full clauses. However, this does not occur until stage 3: in stages 1 and 2 subjects in [-SA] clauses occur at approximately the same rate as subjects in full clauses (17%). Figure 8 below shows the rate of overt subjects in child [-SA] clauses and child full clauses. Notice that the relative increase in subjects in [-SA] clauses is greater than in full clauses.
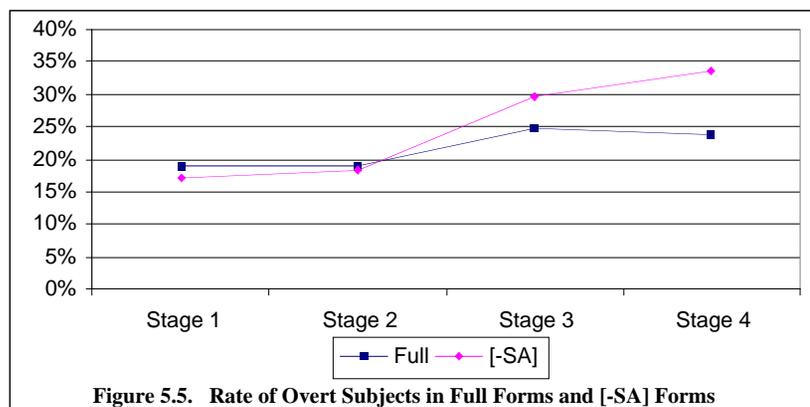

Figure 5.5. Rate of Overt Subjects in Full Forms and [-SA] Forms

Below are some examples of child [-SA] clauses with overt subjects:

(37)   nami   me – ganyang – a   MUS03 line 1560
                   Sameer pr.perf.–beat – IND
   Target =   Sameer   a – me – ni –kanyang–a
                   Sameer SA$_{3s}$–pr.perf.–OA$_{1s}$– beat – IND
                   'Sameer has beaten (me)'
(38)   mimi me – tow – a   MUS05 line 1559
                   I   pr.perf.–remove–IND
   Target =   mimi ni – me – tow – a
                   I   SA$_{1s}$–pr.perf.–remove–IND
                   'I have removed (it)'

(39)   gali na – end – a   MUS08 line 1181
                   car pres–go –IND
   Target =   gari   i – na – end – a
                   Car-5 SA$_5$ – pres – go – IND
                   'The car is going'

(40)   Muko na Charlie mw–ingine na – va – a   vi – atu   FAU02 line 991
          Muko and Charlie 1-other   pres–wear–IND 8–shoe
Target = Muko na Charlie mw-ingine wa – na – va – a   vi–atu
          Muko and Charlie 1-other SA$_{3pl}$–pres–wear–IND   8–shoe
          'Muko and the other Charlie are wearing shoes'

Examples (37) through (40) show [-SA] clauses with different kinds of subjects. The subject of (37) is a name, while in (38) and (39) we find a pronoun and a lexical DP, respectively. In (40) the subject is a conjunction of two names. These are representative of the subjects in [-SA] clauses overall. It should also be noted that while overall quantifiers were rare (a total of 16 tokens), quantifier subjects never occurred in [-SA] clauses in child speech. While this is not conclusive evidence that quantified subjects

are disallowed in [-SA] clauses, it is suggestive that the principles that constrain adult [-SA] clauses operate in child language as well.

Recall, that in [-SA] clauses what look like subjects are in fact topics. So the figures for [-SA] clauses in figure 5.5 shows the rate of *topics* by children while the figures for full clauses show overt *subjects*. Two results emerge from these data. First, we see that at all stages, the proportion of *subjects* in full clauses is close to adult rates (18%, c.f., figure 6). Second, the proportion of *topics* in [-SA] clauses is initially rather low as compared to adult rates, but in stage 3 there is a shift toward the adult norm, and in stage 4 children are using topics at approximately the same rate as adults. If we take the similar frequencies to indicate that the underlying systems are the same (as does Valian, for example), then we see that children at these ages are well-attuned to the constraints on the use of subjects versus topics.

The fact that children by stage 4 show similar rates of topic- and subject-use as adults suggests that children have acquired the syntactic principles governing subject and topic use described earlier in this chapter in section 5.2.3. Specifically, we assume children know that SA identifies a null subject, and the absence of SA (and presence of T) requires a null constant. They also know that this null constant is bound by an anaphoric topic operator, which can be realized as phonologically null or overt, depending on pragmatic restrictions. The most parsimonious account of the parallel behavior of children and adults with respect to the use of subjects in full clauses and [-SA] clauses is that children by stage 3 have an adult-like grammatical system.



Figure 5.6. Adult and Child Subjects and Topic
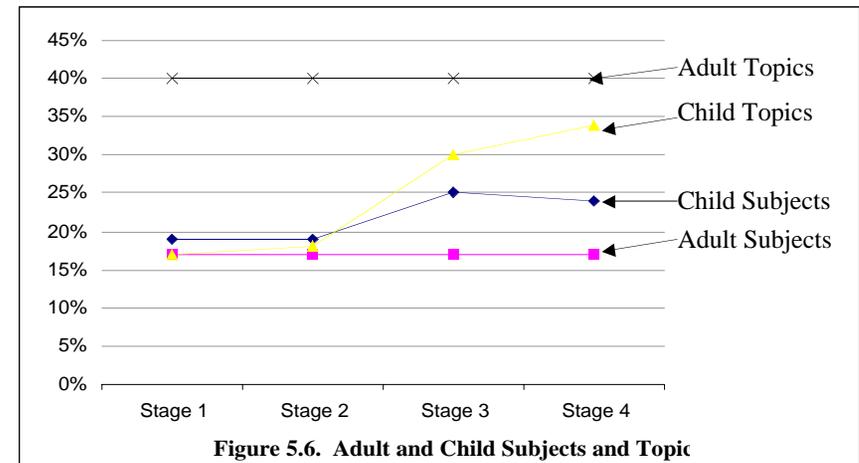
What is especially interesting is that [-SA] clauses are very rare in the adult language (approximately 5% of all sentences), so the properties of the different clause types either are acquired by children with the aid of very rare input or follow from general principles of grammar. In the next two sections we will see that the principles that constrain "subjects" in full clauses and [-SA] clauses also hold in [-T] clauses and bare stems. These latter two clause types are not attested in the adult language, and thus underscore the fact that children are operating according to grammatical principles and not statistical frequency.

5.3.3   [-T] clauses

Recall that [-T] clauses lack a tense marker, but have an overt SA marker. Recall also that approximately 20% of all verbal utterances in the early stages are [-T] clauses, but by stage 3 they decrease to under 5 % and that [-T] clauses are unattested in adults.

What do we expect under the analysis of subjects and topics? According to our analysis, overt subjects are assigned Case by tense. In the absence of a tense specification, there is no possibility for an overt subject. The second option – a topic – is also excluded. Under the assumptions laid out in section 2.14, a topic binds a null constant in subject position. The topic licenses the null constant and also identifies it. Because the null constant does not have any _-features, it cannot check agreement features, and so the null constant is not compatible with SA morphology. Therefore, when SA morphology appears on the verb (as is does in [-T] clauses), it indicates the subject position is filled by something other than a null constant, i.e., either a null expletive or *pro*. But both these options require Case, which is not available in a [-T] clause. Therefore we expect that
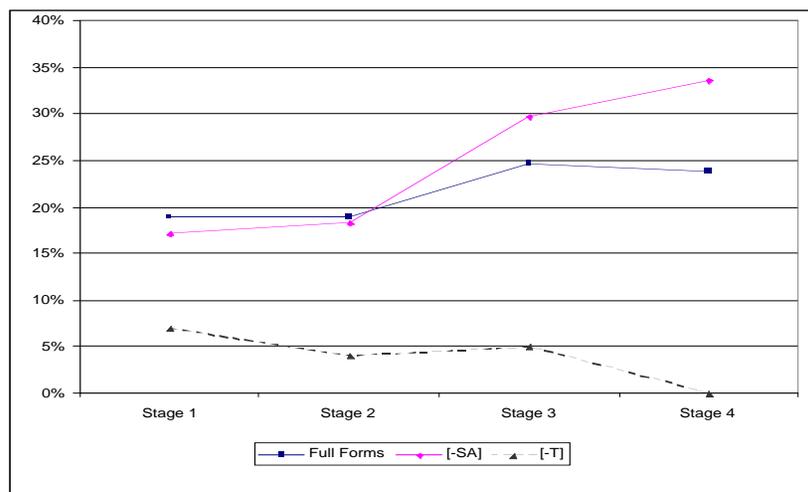
neither subjects nor topics may arise in [-T] contexts.

Looking at the proportion of overt subjects in [-T] clauses across the four stages, we see a striking difference from full clauses and [-SA] clauses. In stage 1, overt subjects occur at a rate of 7%, and this is the

highest rate in [-T] clauses across all stages. Overt subjects decrease to 4% and 5% in stages 2 and 3 respectively, and then fade out entirely by stage 4. The difference between the rate of overt subjects in [-T] clauses and full clauses was found to be significantly different (p=0.01, $\_^2$=19.767, with Yates Correction Factor), as is the difference between overt subject rates in [-T] clauses and [-SA] clauses ((p=0.01, $\_^2$=23.293, with Yates Correction Factor). The percentages and number of tokens are given in table 5.17 below:

Table 5.17 Overt and null subjects in [-T] clauses

Stage	1	2	3	4
Overt Subjects	3 (7%)	1 (4%)	1 (5%)	0 (0%)
Null Subjects	40	24	19	26
	43	25	20	26

In fact, there were only 5 [-T] clauses in the entire corpus that had an overt subject. Below is the exhaustive list of all the [-T] clauses with subjects the children produced[11].

(41)   mimi ni – fungu – a   Haw07, line 387
            I   SA$_{1s}$– open –IND   (Stage 1)
   Target =   mimi ni – na – fung– a
            I   SA$_{1s}$–pres–open–IND
            'I am closing (it)'

(42)   mimi ni – namaz – a   Mus06, line 553
            I   SA$_{1s}$–be quiet–IND   (Stage 1)
   Target =   mimi ni – na – namaz – a
            I   SA$_{1s}$ – pres – be quiet – IND

---

[11] Notice that all the examples were produced by only two of the children. These two children are the least linguistically mature of the four in the sample. There was one example from stage 3, and this example was in the first file in stage three for this child, indicating that perhaps the child was not fully through the stage at the time this recording was done.

'I am being quiet'

(43)	dadi	ni – um – a	Mus08, line 230
	Daddy	SA1s–hurt–IND	(Stage 2)
Target =	Daddy,	ni – ku – um – e	
	Daddy	SA1s–OA2s–hurt–SUBJ	
	'Daddy, I will/want to hurt you'		

(43) dadi ni – um – a Mus08, line 230
Daddy SA$_{1s}$–hurt–IND (Stage 2)
Target = Daddy, ni – ku – um – e
Daddy SA$_{1s}$–OA$_{2s}$–hurt–SUBJ
'Daddy, I will/want to hurt you'

(44) taa iy – ek – a ndani Mus08, line 684
light-9 SA$_9$–be on–IND inside (Stage 2)
Target = Taa i – na – wak – a ndani
Light-9 SA$_9$–pres–be on–IND inside
'The light is on inside'

(45) nn – end – a tuko aa – tembe–a Mus21, line 659
prefix–V– IND Tuko SA – V – IND (Stage 3)
Target = a – na – end – a Tuko, a – ka – tembe – a
SA–pres–go–IND Tuko SA–cont–walk–IND
'He goes (out), Tuko, and then he walks.'

Furthermore, this is a very conservative count. In four of the five examples it is possible that the clause is not a [-T] clause, but something else. In example (42), the verb root is 'namaza', and it is used with an on-going interpretation. An on-going interpretation in Swahili is indicated by the present tense marker [na], as indicated by the target utterance. In Swahili, most verb stems conform to a CVCV pattern, and so there are two possible analyses of the child's '*ninamaza*'. The first is as I have glossed the utterance in (42) above, namely 'ni' is the SA, and 'namaza' is the verb stem. The second possibility is that the child has misanalyzed the verb stem as 'maza' (hence conforming to the general CVCV pattern in Swahili), and has misanalyzed the onset syllable of the root 'na-' as the present tense marker 'na'. Therefore, example (42) could be counted as a misanalyzed full clause instead of a [-T] clause, reducing our number of [-T] clauses with subjects down to four.

Example (43) may also be a candidate for removal from this count. This utterance expresses an intention, and the usual manner for this in Swahili is to use the subjunctive. The adult version would be 'Dadi, nikuume', meaning 'Daddy, I may hurt you'. With the direct object left-dislocated[12], and the verb in the subjunctive, the absence of T is perfectly natural. So instead of this counting as a [-T] clause, it could equally well be counted as a subjunctive clause with a mispronounced final vowel. While such 'errors' in the final vowel are rare, it can be seen from table 5.14 that subjects co-occurring with [-T] clauses are equally rare. Furthermore, context suggests that 'Dadi' is the object of the verb, not the subject, and so it is possible that our overall count of [-T] clauses with subjects is down to three.
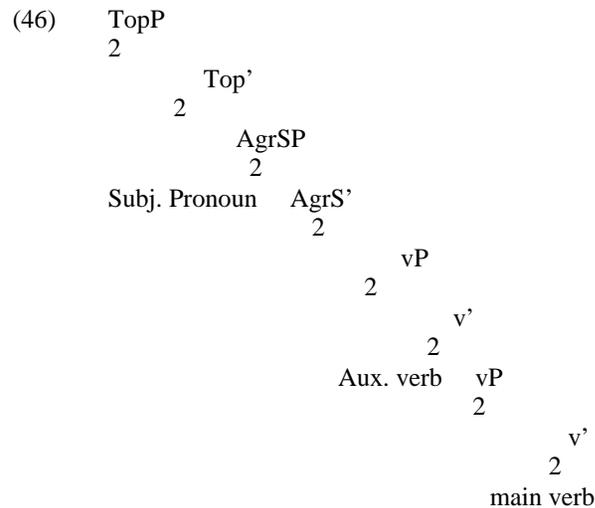
Example (44) is difficult to interpret because the prefixes are not clearly pronounced, and so there is some doubt as to what is being said. The best estimate is as coded. The adult 'target' would be 'inawaka', but with some phonological reduction and assimilation, it is possible that this is not a [-T] clause but again, a full clause. Example (45) is possibly a full clause because the vowel is slightly elongated. In Swahili there is a phonological variant of the present tense marker where [na] alternates with [a]. Since the 3rd person singular SA marker is also [a], this would result in a long [a:]. The lengthening was not adequate in my judgment to code it as a full clause, but this is a possibility.

Each of these utterances was coded and counted as [-T] clause

---

[12] The intonation in this sentence was very clearly 'comma intonation', i.e., the left-dislocated object had raised intonation, and was followed by a pause. I checked this utterance with a second native speaker consultant and we agreed that the intended meaning was with 'Dadi' as the direct object.

because the context and morphology suggest that this is most likely. However, as I have just explained, these five examples are not entirely convincing. My point is simply that we should keep in mind that the count of 5/114 [-T] clauses with subjects is a generous one, and the actual number could well be lower than this.

This result has implications for the analysis put forward in chapter 2 where we argued that SA is true agreement and not a pronoun. Recall that we presented arguments against Zwart (1997) who claims that SA in Swahili is a subject clitic pronoun (not agreement) and tense in Swahili is actually an auxiliary verb. Under Zwart's analysis, all preverbal DPs are topics, not subjects. According to Zwart, the structure of the Swahili clause is as in (46) below where the label 'subject pronoun' corresponds to what I have been calling Subject Agreement (SA), and 'auxiliary verb' corresponds to what I have been calling Tense.

(46)   TopP
         2
                  Top'
              2
                  AgrSP
                  2
         Subj. Pronoun   AgrS'
                      2
                          vP
                      2
                          v'
                      2
                  Aux. verb   vP
                          2
                              v'
                          2
                          main verb

Zwart argues that all preverbal DPs are topics coreferent with the true subject (SA). Recasting our results in Zwart's terms, [-T] clauses are clauses in which the auxiliary verb has been omitted. In such clauses, we found that preverbal DPs (topics, for Zwart) are completely absent. If Zwart is right, the question arises as to why the absence of an auxiliary verb should result in a complete absence of *topics*. Our analysis, on the other

Figure 5.8. Overt Subjects by Clause Type
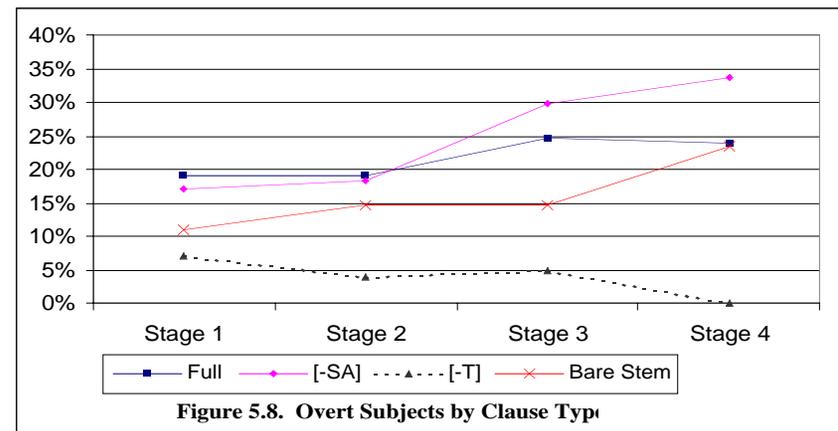
hand, makes a different prediction. Under our analysis, the omission of tense has no effect on *topics*, but it does prohibit the use of true subjects – a natural effect, given that tense assigns case to subjects.

Therefore the lack of tense results in a prohibition on subjects. Moreover, the presence of SA means that there is no null constant in subject position because the null constant has no _-features. The lack of a null constant means that a topic is also lacking in [-T] contexts. Therefore the overall proportion of subjects and topics is very low, as we predicted

151

earlier[13].

5.3.4   Bare Stems

The last clause type we will consider is the Bare Stem. The bare stem has no SA and no T, plus the final mood vowel. This is schematized below in (47) (again, ignoring optional affixes):

(47)   Ø – Ø – V – IND

A generalization that has emerged in several European languages is that null subjects occur predominantly in non-finite contexts (see section 5.1.2 where we discussed the PRO analysis of null subjects). In Swahili bare stems (as we saw is the case in [-T] clauses) we expect no subjects because there is no case assigner. This is not, however, what occurs. Children in early stages of acquisition use subjects in bare stem clauses at a rate of 11% in stage 1, rising to 15% in stages 2 and 3, and then rising further to 24% by stage 4.

As noted above, the lack of tense results in a lack of Case assignment, which prohibits overt subjects. The lack of SA results in the lack of identification, and so *pro* is blocked. The only option is a null constant, which satisfies the EPP. The implication is that the 'subject' of a bare stem is not a subject, but a topic, a topic is required in order to identify the null constant.

5.3.5   Summary

---

[13] These data are consistent with a PRO analysis of [-T] clauses. We rejected PRO as a possibility in the other clause types because overt subjects alternate with null subjects – a characteristic not typical of PRO. However, in the case of [-T] clauses, we have no such alternation: subjects are always null. This suggests that subject position in [-T] clauses contains PRO. See below in the text.

In summary, the picture that emerges from the data on subjects in early Swahili is a complex one. I proposed an analysis of adult [-SA] clauses in which a null constant occurs in subject position and is bound by a topic operator. I proposed that children acquire this property of Swahili early on, and this can be seen in the differential rates of subjects used by children in different clauses. We saw that children use overt "subjects" at roughly adult rates in both full clauses (approximately 18%) and [-SA] clauses (approximately 35%), suggesting that the systems governing the two clause types are the same in children and adults. We also saw that [-T] clauses have virtually no overt subjects. [-T] clauses prohibit true subjects because of the lack of a Case assigner, and a null constant is blocked because of the presence of subject agreement. Therefore, neither a subject nor a topic occurs in [-T] clauses. In bare stems, in contrast, a null constant may occur in subject position because subject agreement is lacking. Therefore the "subjects" that we see in bare stems are topics.

Recall from chapter 2 we discussed all the possibilities for subjects in the various adult Swahili clause types. We concluded that adults allow *pro*, overt subjects, as well as PRO (in the appropriate contexts, i.e., tenseless clauses in which there is either a controller or arbitrary reference is assigned). We concluded that while these three possibilities do arise in full clauses in adult Swahili, [-SA] clauses are different: only the null constant and topic construction is permitted in a [-SA] clause. Recall that *pro* is not possible in a [-SA] clause because it requires identification (which is not possible in the absence of SA), and PRO is not likely because PRO generally does not alternate with overt DPs, whereas subjects in [-SA]

clauses are optionally overt or null. We assume the same holds for child Swahili.

However, what rules out PRO from [-T] clauses and bare stems? Let us consider bare stems first. We saw in section 5.2.4 that Bare stems occur with overt subjects between 11% and 24% of the time in child Swahili. This alternation between overt and null subjects suggests that PRO is not involved as PRO does not alternate with overt DPs. However, [-T] clauses do not alternate with overt DPs, and so PRO is a possibility. In fact, we will assume that in [-T] clauses PRO is in subject position.   Therefore the full inventory of subject options in child Swahili consists of lexical subjects, *pro*, PRO, and null constant + topic, as summarized in table 5.16[14], reproduced below.

Table 5.16 Summary of clause types and "subject" options

	Full Clause	[-SA] Clause	[-T] Clause	Bare Stem
pro	a	*	*	*
Lexical Subject	a	*	*	*
n.c. + Topic	*	a	*	a
PRO	*	*	a	*

Let us turn now to the three theories of subject omission we discussed earlier. We will evaluate each theory in terms of whether it is compatible

---

[14] In this table I have classified bare stems as NOT allowing PRO subjects. In fact, it is possible that some bare stems allow PRO subjects since bare stems are tenseless clauses. Furthermore, since the majority of bare stems have null subjects, it is entirely possible that some of these clauses have PRO subjects. However, it is not possible to determine from this data whether subjects are PRO or null constants. We would need to conduct experimental tests to see whether children allowed, for example, quantified antecedents in bare stems. If they did, this would suggest PRO is a possibility (since PRO can be controlled by quantifiers, e.g., [every child tried [PRO to read the Lord of the Rings]]).

with the subject omission facts of Swahili.

5.4   Processing Revisited

Recall that in section 5.0.1 we considered a proposal by P.Bloom (1990) in which he claims that null subjects occur as a result of processing limitations. Sentences with longer VPs pose a greater processing load and should thus have a higher null subject rate. In this section I investigate Bloom's proposal with respect to Swahili. Bloom only considered processing with respect to non-null subject languages such as English, but I will extend his proposal to Swahili – a null subject language.   In order to do this, I must modify Bloom's proposal in two ways. First, I propose a refinement of Bloom's use of VP length, arguing that it does not accurately capture processing load. I argue that a fuller measure of utterance length is more appropriate (which I call TP length). Second, I argue that measuring processing load in *words* is not suitable for all languages, especially agglutinative languages like Swahili. I thus propose measuring processing load in morphemes. To apply Bloom's hypothesis to Swahili, I investigate two questions: i) whether processing limitations account for subject drop in Swahili full clauses, and ii) whether Bloom's hypothesis can be extended to account for the omission of prefixes in early Swahili. I conclude that processing accounts neither for the omission of subjects in Swahili full clauses nor the omission of SA in early Swahili.

5.4.1   Measuring Processing Load

Bloom assumes two measures of processing load: phonetic content of subjects (i.e., null subjects carry a lower processing load than pronouns, which carry a lower processing load than overt DP subjects) and

VP length in words (i.e., the longer the VP, the greater the processing load).

His procedure for determining VP length is given below (taken from

Bloom, 1990, p.496):

> "VP length was calculated by counting the number of
> words from the verb until the end of the sentence. Proper
> names like *Mommy* and *Daddy* that appeared at the end of
> the sentences were not included in the calculation of VP
> length if it was clear that they were not part of the VP.
> For instance, *I goed to bathroom, Mommy* was counted as
> a VP that is three words long, not four.

This definition of VP length is vague in several respects. For example, in

cases in which words occur between the subject and verb (for example

adverbs or negation) are they included in the calculation of VP length?

(48)   I not goed to bathroom, Mommy

Bloom's description suggests that they are not, but surely negation

contributes to processing load. Bloom does not elaborate on his counting

procedures, but following the logic of his proposal, it is reasonable to

assume that adverbs and negation also contribute to processing load and

should be included in the count of VP length. Therefore let us refine our

procedure for calculating VP length so as to capture this intuition. Bloom's

intention was to calculate processing load, and since subjects were the focus

of his study, subjects could not be included in this calculation. Thus we

can simply calculate processing load as the utterance length excluding the

subject and vocatives. Since this measures more than the length of the *Verb

Phrase*, let us call this TP length[15]:

> Calculate TP length by counting the number of words
> after the subject (if there is one) until the end of the
> sentence. If there is no subject, count all words in the
> sentence. Do not include proper names like *Mommy* and
> *Daddy* that appear at the end of the sentences if it is clear
> that they are not part of the VP.

So, using our definition of TP length, a hypothetical sentence such as (49)

is counted as 4 words long:

(49)   I not goed to bathroom
       1   2   3   4

For English this is an appropriate extension of his measure, but as we will

see below, it requires further refinement for Swahili.

5.4.2   **Utterance length in words/morphemes**

Because Swahili full clauses have a greater phonetic content than

[-SA] clauses, we expect fully specified verbal complexes to occur in

overall shorter utterances. However, measuring utterance length in terms of

words is problematic for agglutinative languages such as Swahili since the

majority of utterances are comprised of a single word: the verbal complex.

Recall that the verbal complex consists minimally of SA-T-V-IND, with

both subject and object optionally null. Since the preferred option is to

have null arguments, most Swahili utterances consist of the verbal complex

---

[15] I do not include postverbal subjects in this analysis for two reasons:
Bloom in his formulation does not include postverbal subjects, and
postverbal subjects are relatively rare in child Swahili.

alone.[16]

Calculating utterance length in words would yield the same result for the two sentences in (51). This is a problem that does not arise in English, but it does in agglutinative languages, and so a recasting of Bloom's measuring criterion in terms of morphemes is required.

(51)   a.   ni – li – kul – a   Swahili.
              SA_{1s}–past– eat–IND   VP length = 1 word
              'I ate'

        b.   ni – li – ki – kul – a   Swahili
              SA_{1s}–past–OA_7–eat–IND   VP length = 1 word
              'I ate it'

Using these revised measures, we can now investigate two possibilities in Swahili. The first is that null subjects in child Swahili are a result of processing limitations. The second is that the omission of SA in Swahili is the analogue of subject omission in English, and hence there should be a relation between SA omission and TP length.

5.4.3   Null Subjects as a result of Processing Limitations

I calculated the length of TP in full clauses with overt subjects full clauses with null subjects. In calculating TP length for full clauses I included all morphemes of the verbal complex (SA, T, OA, V, suffixes, and

---

[16] While the verbal complex is generally thought to be a single prosodic word, there is evidence that it is comprised of two phonological words: [SA+T] [OA-V-IND]. See chapter 2 for discussion of relevant data. For our purposes here, it makes no difference which analysis we assume. If we assume a single-word analysis, our overall VP length will be very close to 1.00. If we assume a two-word analysis, our overall VP length will be close to 2.00. The point remains the same: a calculation based on words for an agglutinative language such as Swahili is meaningless.

Mood), the overt object (if present), negation, adverbs, and demonstratives. I excluded vocatives, as did Bloom, as well as overt subjects. The results are presented in table 5.18 below.

Table 5.18 TP length in morphemes of full clauses with overt and null subjects

	Overt Subject	Null Subjects
Number of clauses	108	446
Number of morphemes	699	2364
Ratio	5.472	5.300

A t-test was performed and the difference in ratio of morphemes-to-clauses was not found to be significant (_=0.01, t=0.175). Therefore subject omission in full clauses in Swahili is not an effect of utterance length. It must be said that this result is completely expected, since the same phenomenon occurs in adult Swahili (and indeed all the other null subject languages of the world). We would hardly want to say that null subjects in adult Italian, for example, occur because of processing limitations. The interesting question is whether processing limitations account for the omission of SA in child Swahili.

5.4.4.   SA Omission as a result of Processing Limitations

If SA omission is the analogue of subject omission in English, then we expect there to be a relationship between SA omission and utterance length. Specifically, we expect [-SA] clauses to occur in longer VPs than full clauses, since full clauses carry a greater processing load. I examined the files of the Swahili children and calculated TP length in [-SA] clauses. We have already established that overt subjects do not affect TP length

significantly, so I included subjects in the count as well.[17] The results are
presented in table 5.19. below:

Table 5.19 Utterance length in [-SA] and full clauses

	[-SA] Clauses	Full Clauses
Number of clauses	486	554
Number of Morphemes	1976	3063
Ratio	4.06	5.52

We see in table 5.19. that full clauses occur in longer utterances
(5.52 morphemes per utterance) than [-SA] clauses (4.06 morphemes per
utterance).[18] This is exactly the opposite of what a processing account
predicts for the omission of SA.

In conclusion, a processing theory of either subject omission or SA
omission finds no support in utterance length. An additional challenge for
the processing account is that [-SA] clauses also occur in adult Swahili. It
is unlikely that adults omit SA due to processing limitations. Furthermore,
to the extent that the proportion of subject omission in [-SA] clauses reflects
that of the adult grammar, this could argue against a processing account of
children's subject omission as well.

---

[17] The ratio of morphemes-to-utterances for [-SA] clauses with null subjects
is not significantly different: 3.96. Thus there is no significant difference
in TP length whether subjects are overt or null.

[18] Full clauses include SA, which by definition is not included in [-SA]
clauses. If we exclude the SA marker in the count of morphemes for full
clauses, then the ratio of morphemes-to-utterances falls to 4.52. Thus full
clauses are longer than [-SA] clauses even after we eliminate SA from the
full clause count.

5.4.5   Child Null Subjects as PRO

The PRO theory of null subjects claims that child null subjects in
non-null subject languages like English and French are like adult non-finite
clauses that contain PRO. Such a theory claims that in child language all
unadult-like null subjects (i.e., non-pro subjects) are PRO. Can this account
for the Swahili null subject phenomenon? The crucial facts to consider are
the following:

Table 5.20   Summary of overt and null subject possibilities in the four clause types

	Overt Subjects	Null Subjects
Full Clause	b	b
[-SA] Clause	b	b
[-T] Clause	*	b
Bare Stem	b	b

Where 'b' indicates that overt or null subjects are attested,
and '*' indicates that they are not attested.

Since adult Swahili is a null subject language, we can safely assume that the
null subjects in full clauses are not PRO, but are in fact *pro* (see chapter 2,
section 2.9). Can the null subjects in the other three clause types be
accounted for under a theory of PRO? PRO rarely alternates with overt
DPs, and occurs exclusively in tenseless environments. Because tense is
specified in [-SA] clauses and null subjects freely alternate with overt DPs,
null subjects in [-SA] clauses cannot be PRO.   The remaining two clause
types are tenseless. If null subjects are PRO in these clauses, they should
both allow null subjects and disallow overt subjects. As mentioned earlier,
[-T] clauses allow null subjects and disallows overt subjects. Therefore [-T]
clauses are compatible with a PRO analysis. However, bare stems alternate
with overt DPs. This suggests that either bare stems constructions do not all
contain PRO subjects (see footnote 14), or that null subjects in bare stems

belong to a different category than PRO.

5.4.6   Topic-Drop

Can these facts be accounted for under a topic-drop analysis? A topic-drop analysis is compatible with the facts of all the clause types except the crucial difference between [-T] clauses and bare stems. Topic drop is unrelated to the inflectional marking that occurs on a verb. In other words, topic drop can occur in a fully specified clause as well as an underspecified clause, provided the structural requirements of topic drop are present (i.e., an available landing position in the left-periphery and c-command of the trace). However, we see here that particular inflectional underspecifications have particular effects on the occurrence of subjects. A topic-drop analysis can account for the different rates of null subjects in full clauses and [-SA] clauses, since topic drop may occur differentially under the different pragmatic conditions that each clause type occurs in. However, it is unclear why overt subjects are completely blocked in [-T] clauses, while they are possible in bare stems. The only difference in these two clause types is whether they are specified for agreement or not, and topic-drop is not sensitive to this specification. Therefore the Swahili subject facts are not fully compatible with a topic-drop analysis.

In this next section I will recapitulate some of the results we have seen so far, and suggest a theory of null subjects in child Swahili that accounts for the facts.

5.5   Summary of the Chapter

In this chapter we discussed subjects in Swahili child language. Several results emerge that shed light on the nature of child language. First, we saw that children use null subjects in roughly the same proportions as adults in full clauses. This is true at all ages and all stages, and is true for all four children in the study. Second, we saw that children mirror adults in using a relatively larger proportion of overt subjects in [-SA] clauses. I argued that [-SA] clauses involve a null constant bound by an anaphoric topic operator, as proposed in chapter 2. The similarities in frequency between adults and children suggests that children are attuned to this characteristic of Swahili. We then investigated the occurrence of subjects in the two clause types that are attested only in child language: [-T] clauses and bare stems. We saw that in [-T] clauses children use virtually no subjects, while in bare stems, surprisingly, overt subjects do occur. This apparent paradox requires an explanation, since it is quite counterintuitive that the presence of additional features in [-T] clauses (i.e., SA features) disallows overt subjects. I argued that in [-T] clauses, the absence of a null constant in the presence of SA disallows topics, and the absence of T disallows subjects.   Hence, the overall absence of "subjects" in [-T] clauses. I argued that bare stems allow the null constant/topic operator construction, hence the occurrence of "subjects" in these constructions. Tables 5.21 and 5.22 below summarize the compatibility of overt subjects and various null elements in Swahili with the two elements of inflection in Swahili. We see that overt subjects require both +SA as well as +T in order to occur, while the topic/null constant construction occurs only when SA is specified as [–SA]. The topic/null constant construction is compatible with either specification of T. *pro* requires +SA and +T in order to occur, while PRO requires –SA and –T in order to occur.

Table 5.21 Possible subjects with the four inflectional possibilities

	Overt subject	Topic/nc	*pro*	PRO
+SA, +T	+	−	+	−
+SA, -T	−	−	−	−
-SA, +T	−	+	−	−
-SA, -T	−	+	−	+

Table 5.22 Summary of compatibility of null elements and inflectional prefixes.

	+T	-T
+SA	Overt subjects *pro*	−
-SA	Topic/nc	Topic/nc PRO

We then argued against the possibility that subject omission in either full clauses or [-SA] clauses is due to processing limitations. We showed that there is no correlation between null subject clauses and VP length (with VP length refined so as to suit its application to Swahili). Furthermore, we showed that SA omission in child Swahili cannot be attributed to processing limitations as there is no correlation between the omission of SA and utterance length.

The overall theory of subjects in child Swahili that I am proposing is based on two independent mechanisms: the licensing of the topic operator/null constant construction by the absence of agreement features, and the prohibition of overt subjects by the absence of tense features. The first mechanism is a simple extension of Rizzi's (1992) proposal, while the second mechanism follows directly from theories of case assignment through tense. It is only in this way that we can capture the seemingly paradoxical difference between overt subjects in [-T] clauses (virtually

absent) and overt subjects in bare stems (present): the overt "subjects" in bare stems are topics. I thus do not require recourse to topic-drop accounts, as the entire theory is based on the presence or absence of inflectional features in the child utterance. It should be noted that this theory of subjects in child Swahili is reminiscent of the proposal we adopted in chapter 4 for the omission of inflectional elements: the Agr-Tense Omission Model of Schütze & Wexler (1996).

Two major conclusions can be drawn from the findings in this chapter. First, the clause types that we described in detail in chapter 4 show systematic correlation with overt and null subjects. Not only do the proportions of subjects in the various clause types occur according to the syntactic theory presented in chapter 2 (i.e., the complete absence of subjects or topics in [-T] clauses), they also conform to adult proportions insofar as they are possible in the adult language (i.e., children's subjects occur in similar proportions for full clauses and [-SA] clauses). The conclusion we draw from this is that the four clause types are not a result of random omission of inflectional prefixes due to processing limitations, the lack of phonetic salience, pronunciation difficulties, etc. Rather, these omissions are the result of syntactic processes that have specific properties. In this regard, these results complement the conclusions we reached in chapter 4, where we discussed the Metrical Omission Model. We concluded that metrical omission does not predict the omission patterns of prefixes, while here we show that processing limitations do not account for the omission of SA nor the omission of subjects in the different clause types.

The second conclusion concerns the acquisition of silent elements

in syntax. We saw that in languages such as English and French, children make use of two possible mechanisms for null subjects: PRO in tenseless clauses and null topics in tensed clauses (cf. Bromberg & Wexler, 1995). We can add to this inventory of null elements that children know from the data in Swahili. First, they use *pro* correctly and in an adult-like way in full clauses. Not only do null subjects in full clauses occur at roughly the same frequency as in the adult language, there are no restrictions on the reference of the null subject. Second, Swahili children show knowledge of the adult construction of [-SA] clauses which involves a null constant and an optionally null topic operator. Not only do they adhere to the principles of [-SA] clauses, they extend the contexts in which the null constant may occur to bare stems – a construction that they never hear in root context. They also show knowledge of the fact that it is not the absence of Tense that allows the null constant/topic operator construction, but rather the absence of SA. These types of intricate patterns in syntactic constructions in child language argue against a distributional learning algorithm for learning syntax, as it is unclear how children would know the properties of the null constant/topic operator construction, given that there is no negative evidence that could trigger this. It is also unclear how children know that null constants do not occur in [-T] clauses but do occur in bare stems.

Finally, how does this fit into the classification of languages that I outlined in chapter 1? Recall from chapter 1 that I described a classification of child languages into three groups: RI languages (like German, Dutch, French, etc.), non-RI languages (like Italian, Spanish, Catalan, etc.) , and bare stem languages (like English, Sesotho, Quechua, etc.). I argued that assimilating bare stems to RIs is a mistake since it is now clear that English is not the only child language that allows bare stems. It is clear that Swahili

is a bare verb language and not an RI language. Furthermore, it is clear that Swahili is not like Italian or Spanish in that children acquiring Swahili omit morphology at a relatively high rate.

5.6 Conclusion and Future Research

This study represents the first study of the acquisition of Swahili. We have seen that there are significant differences between the development of Swahili and other Bantu languages such as Sesotho and Siswati. Swahili has the characteristic of marking SA and T as independent prefixes that are obligatory in most contexts, while T may be omitted in Sesotho and Siswati. Thus we were able to compare the emergence of agreement and tense in a single language: something that has not been done before. This study is also one of a handful of studies that investigate null/overt subjects in a null subject language.

In chapter 1, I presented a simplified typology of languages based on the patterns of inflectional elements that children in various languages exhibit. I suggested that RI languages and Bare Verb languages be considered separately, whether the two phenomena are underlyingly related or not. The intentions was not to propose that RIs and bare verbs are distinct in their source, but rather that the surface differences between these phenomena must be clearly distinguished and characterized. The first step in such a process is to group the languages and phenomena based on surface similarities and distinctions, which is what my typology is a first attempt at.

In Chapter 2 I described some social and phonological facts of Swahili, and described the morphosyntax of Nairobi Swahili. I tried to distinguish Nairobi Swahili from Modern Standard Kiswahili (*Kiswahili Sanifu*) whenever necessary, since the children in this study spoke the

former.

I discussed a phenomenon in adult Nairobi Swahili that has not received any attention in the theoretical Bantu literature to date: the omission of subject agreement. Adults in Nairobi Swahili omit subject agreement in a limited set of contexts and under strict syntactic constraints. These constraints suggest that preverbal DPs in so-called [-SA] clauses are not subjects (as they are in full clauses), but rather are topics. I proposed that in [-SA] clauses, the preverbal DP binds a null constant in subject position, which does not license agreement morphology (hence no SA). I showed that this occurs in approximately 5% of all adult verbal utterances – a small but significant proportion.

Chapter 3 described the methodology I employed, including descriptions of the children, the data collection procedures, the transcription format / procedures, and the analysis techniques. In chapter 4 I presented results showing that children omit both obligatory suffixes to varying degrees. I showed that at early stages children sometimes omit only SA, sometimes omit only T, and sometimes omit both SA and T. At the same time, children also produce fully inflected, adult-like clauses which contain both SA and T, suggesting that the omission is not due to the lack of knowledge of the inflectional morphology. I showed that a processing account of inflectional omission (e.g., P.Bloom, 1990) does not satisfy the data, nor does a Metrical account of inflection omission (e.g., Gerken, 1991) or a truncation account of omission (e.g., Rizzi, 1994). I concluded that the data support the Agreement – Tense Omission Model (ATOM) of Schütze & Wexler (1996) and Schütze (1997).

Chapter 5 investigated the occurrence of null and overt subjects in child Swahili. Because Swahili is a null subject language, this is a somewhat complicated issue to investigate. I found that children allow null subjects at approximately the same frequencies as adults. Moreover, when the data are broken down into clause type, children allow null subjects at the same frequencies as adults do in the various clause types. In other words, both adults and children allow overt subjects in full clauses at approximately 17%, and both adults and children allow overt subjects in [-SA] clauses at the rate of approximately 35%. Our tentative interpretation of this fact was that children have acquired the topic-null constant structure that I proposed for adult [-SA] clauses. This conclusion was strengthened when the rate of overt subjects in other child clause types was examined. In [-T] clauses, overt subjects are entirely prohibited. This result follows naturally from a theory of PRO – because T assigns case, all overt subject are prohibited. Furthermore, because a null constant cannot license agreement morphology, it is not compatible with a [-T] clause (which contains SA). Additionally, we saw that children allow overt subjects in bare stems, and I argued that these are overt topics binding a null constant. My conclusion was that children at these young ages (approximately age 2) have acquired the properties of the various clause types, and hence have acquired the properties associated with the various syntactic heads.

While these findings are theoretically informative, there are many issues that remain to be resolved. Further investigation is necessary to determine the developmental path of object agreement. As I mentioned in section 4.4.2, children appear to have knowledge of the restrictions on OA fairly early on. In fact, there appears to be a distinct difference between the acquisition of SA and the acquisition of OA. However, as we noted earlier, due to the semantic restrictions on OA, it is not possible to definitively conclude that OA is acquired by the children in this naturalistic corpus.

Experimental data are required in order to determine this.

There has been a growing body of work that points to there being a relationship between inherent lexical aspect (*aktionsart*) and inflectional morphology in child language. Future research will address this issue with regards to these Swahili children: do omissions occur on certain types of predicates?

Additionally, because of the data come from a naturalistic corpus, there were several things that could not be studied. The children tended not to use syntactic negation, and so the complex negation morphology that we saw in section 2.4.3 was almost never exhibited by the children. Because negation and SA are fused into a single morpheme, it would be interesting to see whether [-SA] clauses occur in negative contexts. Similarly, it would be interesting to see if children in early stages produce the correct negative final vowel in the appropriate contexts (recall from section 4.6 that children in the early stages rarely used final vowels other than the indicative). Other questions include whether children use mood correctly at early stages or do they not have knowledge of when to use subjunctive; do children truly obey the syntactic restrictions on [-SA] clauses (e.g., no quantified subjects) or is this simply a product of naturalistic discourse? All these questions involve some measure of experimental elicitation, which was not part of the methodology of this study.

Overall, I think this study fills some interesting empirical and theoretical gaps. With further experimental and naturalistic data collection and analysis, I hope to provide further results that will add to our growing body of knowledge in the field of language acquisition.