

UNIVERSITY OF CALIFORNIA

Los Angeles

The Acquisition of Nairobi Swahili:
The Morphosyntax of Inflectional Prefixes and Subjects

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Applied Linguistics

By

Kamil Ud Deen Salah Ud Deen

2002

The dissertation of Kamil Ud Deen Salah Ud Deen is approved.

Carson T. Schütze

Dominique Sportiche

Thomas Hinnebusch

Alison Bailey

Nina Hyams, Committee Chair

University of California, Los Angeles

2002

List of Tables.....	v	The Syntax of Swahili.....	32
List of Figures.....	vii	2.5 Subject Agreement versus Subject Pronoun	32
List of Abbreviations and Glosses.....	vii	2.5.1 Keach (1995).....	34
Acknowledgments.....	viii	2.5.2 Additional arguments for agreement.....	36
Vita.....	x	2.6 Tense versus Auxiliary Verb.....	38
Abstract.....	xi	2.7 Object Agreement versus Incorporated Pronoun.....	42
Chapter 1.		2.8 The Tensed Clause.....	46
Introduction.....	1	2.9 Null Subjects – null pro.....	49
1.1 The study.....	1	2.10 Subject Agreement Omission.....	52
1.2 Early Morphosyntactic Convergence.....	2	2.10.1 Imperatives.....	53
1.2.1 The Frequency of Omission.....	2	2.10.2 Infinitives.....	53
1.2.2 The Paucity of Agreement Errors.....	3	2.10.3 Habituals.....	54
1.2.3 Root Infinitives.....	4	2.10.4 Continuative Clauses.....	54
1.3 Acquisition of Bantu Languages.....	8	2.10.5 Differences between Habituals and	
1.3.1 Sesotho.....	8	Continuatives.....	54
1.3.2 Siswati.....	10	2.10.5.1 Optionality of SA.....	55
1.4 Other Agglutinative Languages.....	11	2.10.5.2 Optionality of subjects.....	55
1.4.1 Quechua.....	12	2.10.5.3 Embedding.....	55
1.4.2 Inuktitut.....	13	2.11 [-SA] Clauses.....	57
1.5 Organization of the Dissertation.....	15	2.11.1 Frequency of [-SA] Clauses.....	57
Chapter 2.		2.11.2 Tense in [-SA] Clauses.....	58
Morphosyntax of Swahili.....	16	2.11.3 Implicit Reference of [-SA] Clauses.....	58
Introduction.....	16	2.11.4 Overt Subjects in [-SA] Clauses.....	58
2.1 Basic Facts.....	17	2.12 Null Subjects in [-SA] Clauses.....	59
2.2 Phonology.....	18	2.13 Rizzi’s Null Constant.....	60
2.3 Noun Classes.....	20	2.14 Null Constants in Swahili.....	62
2.4 Verbal Complex.....	21	2.14.1 Accounting for [-SA] Characteristics.....	62
2.4.1 Subject Agreement.....	21	2.14.2 Identification in [-SA] Clauses.....	63
2.4.1.1 Nairobi Swahili versus		2.14.3 Ariel (1990) and Gutman (1999).....	65
Standard Swahili.....	22	2.15 Conclusion.....	69
2.4.1.2 SA Omission.....	23	Chapter 3.	
2.4.2 Tense.....	24	Methodology and Data.....	70
2.4.3 Negation.....	25	3.0 Swahili data.....	70
2.4.4 OA Marking.....	26	3.1 Children.....	70
2.4.5 Roots.....	27	3.2 Data Collection.....	71
2.4.6 Suffixes.....	28	3.3 Linguistic Measures.....	73
2.4.7 Final Vowel.....	30	3.4 Staging Criteria.....	73
2.4.8 Imperatives.....	31	Appendix 3A Adult Similarities.....	78
		Appendix 3B Codes used in Transcripts.....	80

Appendix 3C	Staging Data for each Child.....	82
Chapter 4.		
Omission of Prefixes.....		85
4.1	Introduction.....	85
4.2	Theories of language acquisition.....	86
4.2.1	Metrical Omission Model (Gerken 1991)	86
4.2.2	Truncation (Rizzi 1994).....	89
4.2.3	Underspecification of Tense (Wexler 1994)..	90
4.2.4	Underspecification of Agr (Clahsen <i>et al.</i> 1996.....)	90
4.2.5	ATOM (Schütze & Wexler, 1996; Schütze 1997)....	92
4.3	Predictions for Swahili.....	93
4.3.1	Metrical Omission Model.....	93
4.3.2	Truncation Hypothesis.....	95
4.3.3	Underspecification Theories.....	97
4.4	Results.....	99
4.4.1	Attested Clause Types in child Swahili – Overview.....	99
4.4.2	A note about Object Agreement.....	102
4.4.3	Full Clauses.....	104
4.4.4	[-SA] Clauses.....	105
4.4.5	[-T] Clauses.....	107
4.4.6	Bare Stems.....	110
4.4.7	Tense Omission versus SA omission.....	112
4.4.8	Summary.....	113
4.5	Revisiting Acquisition Theories.....	114
4.5.1	Metrical Omission Model.....	114
4.5.2	Truncation Hypothesis.....	115
4.5.3	Underspecification of T.....	115
4.5.4.	Underspecification of Agr.....	115
4.5.5	ATOM.....	116
4.6	Remaining questions.....	116
4.7	Conclusion.....	118
Appendix 4A	Missing Mood Vowel Utterances.....	119
Appendix 4B	Root Infinitives.....	120
Appendix 4C	Monthly numbers of omissions by child	122
Appendix 4D	Subjunctives and their Interpretation.....	123

Chapter 5.

Subject Omission and [-SA] clauses.....	126	
5.0	Introduction.....	127
5.1	Null Subject Accounts.....	128
5.1.1	Processing Limitations.....	128
5.1.2	Null subjects as PRO.....	129
5.1.3	Topic-drop.....	131
5.1.4	Valian (1991).....	134
5.1.5	Wang <i>et al.</i> (1992).....	135
5.2	Adult Swahili.....	137
5.2.1	Swahili is a null subject language.....	138
5.2.2	[-SA] clauses for adults.....	140
5.2.3	Other clause types.....	142
5.2.4	Subjects in different adult clause types.....	144
5.2.5	Topics or Subjects.....	144
5.3	Subject use by children.....	145
5.3.1	Full clauses.....	146
5.3.2	[-SA] clauses.....	147
5.3.3	[-T] clauses.....	148
5.3.4	Bare Stems.....	152
5.3.5	Summary.....	152
5.4	Processing Revisited.....	153
5.4.1	Measuring Processing Load.....	153
5.4.2	Utterance length in words/morphemes.....	154
5.4.3	Null Subjects as a result of Processing Limitations....	155
5.4.4	SA Omission as a result of Processing Limitations....	155
5.4.5	Child Null Subjects as PRO.....	156
5.4.6	Topic-Drop.....	157
5.5	Summary of the Chapter.....	157
5.6	Conclusion and Future Research.....	159

References.....	162
------------------------	------------

List of Tables

Chapter 1

Table 1.1	Summary of languages that allow RIs, disallow RIs and allow bare verbs.....	1
Table 1.2	Proportion of bare verbs in child English (Sano & Hyams, 1994).....	3
Table 1.3	Agreement errors in various languages (Sano & Hyams, 1994).....	3
Table 1.4	Frequency of verbs in first person singular contexts (Harris & Wexler, 1996)	3
Table 1.5	Frequencies of RIs in French and Swedish (Sano & Hyams, 1994).....	4
Table 1.6	RIs and finite verbs in the files of Simone (Behrens, 1993).....	5
Table 1.7	Finiteness versus verb position in German (Poeppel & Wexler, 1993)	5
Table 1.8	Finiteness versus verb position in French (Pierce, 1989).....	6
Table 1.9	Null subjects in finite and non-finite clauses in several child languages (Rasetti, 2000).....	6
Table 1.10	Proportion of RIs in some non-RI languages (Sano & Hyams, 1994)	7
Table 1.11	Inflection dropping by Inuit children and adults.....	13
Table 1.12	Summary of languages that allow RIs, disallow RIs and allow bare verbs.....	14

Chapter 2

Table 2.1	Meinhoff's Noun Classification System.....	21
Table 2.2	Standard Swahili and Nairobi Swahili SA morphology.....	23
Table 2.3	Tense/aspect morphemes in Swahili.....	24
Table 2.4	Swahili Negation Paradigm.....	25
Table 2.5	Standard and Nairobi Swahili Object Agreement Paradigm for noun classes.....	26
Table 2.6	Personal pronoun Object Agreement.....	26
Table 2.7	Some root structures for Swahili verb roots.....	27
Table 2.8	Monosyllabic verb roots and their citation forms.....	27
Table 2.9	Common Verbal Suffixes in Swahili.....	28

Table 2.10	Types of Applicative sentences in Swahili.....	29
Table 2.11	Contexts for use of final vowel.....	30
Table 2.12	Summary of characteristics of habitual and continuative clauses.....	56
Table 2.13	Proportions of different clause types in adult Swahili...	57
Table 2.14	Person reference in [-SA] clauses with Overt and Null Subjects.....	68

Chapter 3.

Table 3.1	Age ranges and number of recordings for each child.....	70
Table 3.2	Codes used in example (1)	72
Table 3.3	Starting and Ending MLUs for all four children.....	73
Table 3.4	Starting and Ending V ratios for the four children.....	73
Table 3.5	Development of MPHs across time.....	74
Table 3.6	Proportion of MPHs on verbs.....	76
Table 3.7	Criteria for stages.....	76
Table 3.8	Criteria for assigning scores for the purpose of staging children.....	77
Table 3.9	Composite scores for the purpose of staging children.....	77
Table 3.10	Composite scores for sections of Mustafa's corpus.....	77
Table 3.11	Division of the Swahili Corpus by stage.....	77

Appendix 3A

Table 3A.1	Adult statistics from various files.....	78
Table 3A.2	Correlation Coefficients for adults.....	79

Appendix 3B

Table 3B.1	Table of codes.....	80
------------	---------------------	----

Appendix 3C

Table 3C.1	Age, MLU, V Ratio and %MPH for Hawa.....	84
Table 3C.2	Age, MLU, V Ratio and %MPH for Mustafa.....	84
Table 3C.3	Age, MLU, V Ratio and %MPH for Fauzia.....	84
Table 3C.4	Age, MLU, V Ratio and %MPH for Hassan.....	84

Chapter 4.

Table 1.12	Summary of languages that allow RIs, disallow RIs and allow bare verbs.....	85
Table 4.1	Summary of possible INFL features for ATOM.....	93
Table 4.2	Summary of Predictions.....	98
Table 4.3	Clause types in Hawa's files.....	100
Table 4.4	Clause types in Mustafa's files.....	100
Table 4.5	Clause types in Fauzia's files.....	100
Table 4.6	Clause types in Hassan's files.....	100
Table 4.7	Frequencies of clause types, by stage.....	101
Table 4.8	Proportions of clause types, by stage.....	101
Table 4.9	Number of OA in each child's files.....	102
Table 4.10	Total number of transitive clauses.....	102
Table 4.11	Total number of intransitive clauses.....	102
Table 4.12	Types/tokens of verbs expressing irrealis mood.....	116
Table 4.13	Proportion of subjunctives to indicative verbs in early stages and late	117

Appendix 4C

Table 4C.1	Monthly data for Hawa.....	122
Table 4C.2	Monthly data for Mustafa.....	122
Table 4C.3	Monthly data for Fauzia.....	122
Table 4C.4	Monthly data for Hassan.....	122

Appendix 4D

Table 4D.1	Total indicative, subjunctive and negative (final vowel) utterances by stage.....	123
Table 4D.2	Subjunctives by stage.....	123
Table 4D.3	Interpretation of children's indicative verbs.....	123
Table 4D.4	Interpretation of bare indicative stems.....	124
Table 4D.5	Interpretation of subjunctive verbs.....	124
Table 4D.6	Summary of Swahili mood marking.....	125
Table 4D.7	Form-meaning contingencies.....	125

Chapter 5

Table 5.1	Omission in obligatory contexts of subjects and objects (P. Bloom, 1991)	129
Table 5.2	Null and Overt subjects in non-finite contexts (Phillips, 1995).....	130

Table 5.3	Null subjects in inflected <i>be</i> contexts (Sano & Hyams, 1994)	130
Table 5.4	Overall proportion of Null subjects (Sano & Hyams, 1994).....	130
Table 5.5	Null and Overt subjects in finite contexts (Phillips, 1995).....	131
Table 5.6	Proportion of null subjects in declarative and <i>wh-</i> utterances at early stages from Adam and Eve (Bromberg & Wexler, 1995).....	132
Table 5.7	Null and pronominal subjects in finite/non-finite <i>wh-</i> questions for Adam (Bromberg & Wexler, 1995).....	132
Table 5.8	Null and pronominal subjects in finite/non-finite <i>wh-</i> questions for Eve (Bromberg & Wexler, 1995).....	132
Table 5.9	Proportion of null subjects in declaratives and <i>wh-</i> utterances at later stages from Adam and Eve (Bromberg & Wexler, 1995).....	133
Table 5.10	MLU, age, proportion of verbs and proportion of overt subjects for Valian's American children (Valian, 1991).....	134
Table 5.11	Proportion of verbal utterances and overt subjects for Valian's Italian children (Valian, 1991).....	135
Table 5.12	Chinese and American Children's proportions of null arguments.....	136
Table 5.13	Overt and Null subjects in the speech of adults.....	139
Table 5.14	Proportions of different clause types in adult Swahili...	143
Table 5.15	Overt subjects in [-T] clauses.....	146
Table 5.16	Summary of clause types and "subject" options.....	146
Table 5.17	Overt and null subjects in child and adult Swahili.....	150
Table 5.18	TP length in morphemes of full clauses with overt and null subjects.....	156
Table 5.19	Utterance length in [-SA] and full clauses.....	156
Table 5.20	Summary of overt and null subject possibilities in the four clause types.....	157
Table 5.21	Possible subjects with the four inflectional possibilities.....	158
Table 5.22	Summary of compatibility of null elements and inflectional prefixes.....	158

List of Figures

Chapter 2	
Figure 2.1	Clause types in Adult Swahili..... 57
Figure 2.2	Overt subjects in adult full clauses and [-SA] clauses... 59
Chapter 3	
Figure 3.1	Development of MPHs by age (developed from Veneziano & Sinclair, 2000) 75
Appendix 3C	
Figure 3C.1	MLU by child..... 82
Figure 3C.2	V Ratio by child..... 83
Figure 3C.3	%MPH by child..... 83
Chapter 4	
Figure 4.1	Overall proportions of clause types for each child..... 100
Figure 4.2	Proportions of the four clause types in child stages and adults..... 102
Figure 4.3	OA in transitive and intransitive clauses in Child Swahili..... 103
Figure 4.4	OA in transitive and intransitive clauses in Adult Swahili..... 103
Figure 4.5	Tense Omission and SA Omission across stages..... 113
Figure 4.6	Overall frequency of subjunctives in early and late stages..... 117
Appendix 4D	
Figure 4D.1	Subjunctives as a proportion of non-imperative verbs... 123
Chapter 5	
Figure 5.1	Null subject proportions in Chinese children, American children and Chinese adults (from Wang <i>et al.</i> 1992)..... 136
Figure 5.2	Overt subjects in Swahili adult full clauses and [-SA] clauses. 144
Figure 5.3	Overt subjects in child and adult Swahili..... 146
Figure 5.4	Subjects in child full clauses and all adult verbal utterances..... 147
Figure 5.5	Overt Subjects in Full Forms and [-SA] Forms..... 147
Figure 5.6	Adult and child subjects and topics..... 148

Figure 5.7	Overt subjects in child Full clauses, [-SA] clauses and [-T] clauses..... 149
Figure 5.8	Overt subjects by Clause type..... 152

List of Abbreviations and Glosses

Appl.....	Applicative suffix
Cond.....	Conditional Tense prefix
Cont.....	Continuative Tense prefix
Fut.....	Future Tense Prefix
Hab.....	Habitual Tense prefix
Ind.....	Indicative final vowel
Inf.....	Infinitive prefix
Loc.....	Locative noun suffix
nc.....	Null constant
NEG.....	Negative final vowel
OA.....	Object Agreement prefix
OP.....	Operator
Past.....	Past Tense prefix
Perf.comp.....	Perfect Completive Tense prefix
Pr.prf.....	Present perfect Tense prefix
Pres.....	Present Tense prefix
Recip.....	Reciprocal suffix
SA.....	Subject Agreement prefix
State.....	Stative suffix
Subj.....	Subjunctive final vowel
T.....	Tense prefix
Full Clause.....	A verbal complex containing at least SA and T
[-SA] Clause.....	A verbal complex in which only SA is missing
[-T] Clause.....	A verbal complex in which only T is missing
Bare Stem.....	A verbal complex in which both SA and T are missing

ACKNOWLEDGEMENTS

Thanks first to Nina Hyams, my advisor and dissertation committee chair. She taught me to think clearly, to look at data impartially, to make sense out of the chaos. Nina worked tirelessly with me from the inception of this project to completion. She helped in all aspects of my degree, including securing funding, data collection / analysis, preparing for job interviews, making sense of an incomprehensible world, and finally the writing of the dissertation. Nina, your vision and passion for your work inspire me in my own work, and for that and all else that you have given to me, I am forever indebted to you, thank you.

The rest of my committee were incredible in their willingness to give of their time and knowledge. Carson Schütze undoubtedly made both my dissertation as well as my research considerably better. Carson is a superb scholar, a fantastic teacher, and a good person. I am extremely grateful for his time, energy, and patience. Dominique Sportiche makes the most complex things seem like child's play. Dominique helped immensely with my qualifying paper on Swahili syntax, as well as every aspect of the dissertation. He took a genuine interest in the acquisition data, and my bumbling ideas about them. For hours in his office, we spoke about syntax, Bantu languages, and other fun stuff. I gained more than I could have hoped from my time with him.

Thanks to Mwalimu Hinnebusch, for his help in the Swahili data, the description of the various dialectal differences, and for hours of time with him in his office discussing things from up-country Swahili to the worth of non-native intuitions. Umenisaidia sana, asante Mzee. Alison Bailey, my outside committee member, took interest in my project from

early on and worked much like an inside member. She gave of her time, and encouraged at all points in the dissertation. Thanks very much Alison, for the help as well as friendship.

Other than my committee, there are countless people who must be thanked and acknowledged. Katherine Demuth, who gave of her time for discussions and sent me many readings in Bantu linguistics and the acquisition of Sesotho – thank you very much. Anoop Mahajan was an undergraduate advisor in the Linguistics department when I was an undergraduate at UCLA, and he opened my eyes to the elegance of syntax and typology. He has always been there for me when I needed help, and has always encouraged me. I suppose I would never have stuck with this if it wasn't for Anoop. Similarly, Anna Szabolci encouraged me as an undergraduate and made me believe that Linguistics was something that I could do the rest of my life. I hope our paths cross once again soon. And Susie Curtiss, for being the most thoughtful, kind and moral person I know at UCLA.

Lyn Repath-Martos, the graduate advisor in Applied Linguistics, is a miracle-maker and an angel. Lyn did everything for me, as I know she does for all the students in Applied Linguistics. She is on top of all problems, she can figure anything out, and she NEVER slips up. Lyn, whatever you get paid, it isn't enough. Thanks Lyn, none of this would be possible without you.

I learned and benefited greatly from the faculty in Applied Linguistics. Thanks to Marianne Celce-Murcia, who helped me with teaching, grammar description, and guided me in various way. Thanks very much to John Schumann for help in many matters, as well as Roger Andersen for work on tense and aspect, and the rest of the Applied

Linguistics faculty. Thanks very much to the faculty in Linguistics, who in all aspects thought of me as one of their students. Tim Stowell, who is the most generous and helpful genius I know, taught me a great deal about syntax, tense, and how to enjoy both. Thanks also to Colin Wilson, Hilda Koopman, Ed Keenan, Russ Schuh, and Bruce Hayes. A special thanks to Pam Munro for all the help over the years.

Thanks to all the friends who helped and encouraged me: Leston Buell, Harold Torrence, Shabnam Shademan, Keiko Okada, Rose Letsholo, Jason Kandybowicz, Haiyong Liu, Adam Albright, Greg Kobele, Stephen Wilson, Masangu Matondo, David Olsher, Leah Wingard, Nathan Carr, Sang Shin, Yuri Kusuyama, Motoko Ueyama, Chris LaBelle, Shigeko Sekine (Keep FLARE alive!), Kathy Howard, Leslie Moore, Nancy Jones, Lexie Wood. I am especially indebted to those in the psycholinguistics lab for camaraderie and patient attendance of various practice talks: Julia Berger-Morales, Manola Salustri (the deadly duo), Ivano Caponigro, Jill Gilkerson, Shannon Madsen, Gianluca Storto, Lisa Brunetti, and Leonie Cornips (visiting postdoc in winter 2002). To those who came before me and helped shape my ideas, thanks very much: Tetsuya Sano, John Grinstead, Misha Becker, and Stella Debode.

I learned to teach from the best of the best: Linda Jensen, Christine Holten, Janet Goodwin, and Donna Brinton. Linda broke me in, Christine fine-tuned me, and with Janet it was all fun. Donna, my guardian angel, shaped my teaching career with great care and forethought. Thanks to all four of you.

My friends Zein Shatry (for endless Swahili judgments and tireless movie reviews), Qaesar El-Haj (toughing it out in Kiambu), Hashim Shariff Shatry (save me some viazi vya madonge), and Alex Nsimbi (I wrote the

biggest essay of them all), made the whole thing easier. And to my dear friend Ahmad Abu-Akel – you deserve a better friend than me, but you stuck by me.

Putting aside the great upbringing and the patience my parents showed with me, they contributed directly to this dissertation in innumerable ways. Their constant love and concern urged me forward, and made me feel as if I could not fail. You know you have a place to stay if you get cold in the winter (there, it's in writing).

And finally, my wife. This is a monumental achievement for me, but it belongs to you. Your courage and direction brought us to this point, I hope you will allow yourself that. You put up with more than one person should be asked to put up with, and somehow you did it with a huge smile on your face. Did you ever think we would actually get here? Thanks for the support, the encouragement, the space when I needed it and the attention when I needed it. I know for certain I would not be here today were it not for you. Let's surf together into the sunset and live happily ever after. Alhamdulillah.

VITA

March 07, 1972 Born, Nairobi, Kenya

1994 B.A. Linguistics
University of California, Los Angeles

1995 Ed.M. Human Development & Psychology
Harvard University
Cambridge, Massachusetts

1995-2000 Teaching Assistant
ESL Service Courses,
University of California, Los Angeles.

1995-2000 Teaching Assistant
Department of Linguistics
University of California, Los Angeles

1997 Pilot Fieldwork, Nairobi, Kenya, under Ford Foundation pre-dissertation fieldwork fellowship

2000-2001 Dissertation Fieldwork, Nairobi, Kenya

2001-2002 Dissertation Year Fellowship
University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Deen, Kamil Ud & Gaby Solomon (1996). ‘Advantages of Computerized paced & timed readings for reading rate development.’ Talk at the USC ESL conference, February 1996.

Deen, Kamil Ud (2001) ‘Swahili verbal prefixes in Child Language.’ Paper presented at the UCLA Humanities Symposium, April 2001.

Deen, Kamil Ud (2001) “The Acquisition of Swahili verbal morphology”. In the Proceedings of Generative Approaches to Language Acquisition. Conference held in Palmela, Portugal, September 2001.

Deen, Kamil Ud & Nina Hyams (2002) “The Form and Interpretation of Finite and non-Finite verbs in Swahili”. In the proceedings of the Boston University Child Language Development Conference held in Boston, MA, November 2001.

Schütze, C. & Kamil Ud Deen (2002). ‘Early inflection: New Evidence for ATOM.’ Talk at University of Maryland, College Park, March, 2002.

ABSTRACT OF THE DISSERTATION

The Acquisition of Nairobi Swahili:
The Morphosyntax of Inflectional Prefixes and Subjects

by

Kamil Ud Deen Salah Ud Deen
Doctor of Philosophy in Applied Linguistics
University of California, Los Angeles, 2002
Professor Nina Hyams, Chair

This study investigates the acquisition of inflectional prefixes in Swahili, an eastern Bantu language. The order of morphemes in adult Swahili is: Subject Agreement – Tense – (Object Agreement) – Verb Root – (derivational suffixes) – Mood Vowel. I present data from an original corpus of 4 Swahili-speaking children (ages 1;8-3;0) who were recorded in Nairobi, Kenya. An analysis of the children’s verbal utterances reveals that four clause types occur in the speech of all four children, with omissions diminishing with maturity:

- | | | |
|----|---------------------|----------------|
| a. | Agr – T – Verb Stem | Full Clause |
| b. | Ø – T – Verb Stem | [-SA] Clause |
| c. | Agr – Ø – Verb Stem | [-T] Clause |
| d. | Ø – Ø – Verb Stem | Bare Verb Stem |

Of these four, only full clauses and [-SA] clauses are permitted by adults in this non-standard dialect of Swahili (Deen, 2002). Furthermore, tense becomes obligatory earlier than subject agreement, the omission of which persists until the latest data points. The data support the Agr-Tense Omission Model (Schütze & Wexler, 1996) in showing that agreement and tense may be optionally and independently underspecified.

Interestingly, the omission of Agr and T has effects on the occurrence of overt subjects, suggesting that the omission is not purely phonological, but rather is of a syntactic nature. When full clauses occur, children allow overt subjects at approximately adult rates (Swahili being a null subject language, this rate is approximately 17%). In [-SA] clauses, overt subjects occur at significantly higher rates in both child and adult Swahili (~40%). In [-T] clauses, overt subjects are entirely unattested. This is expected if we assume that in the absence of T, children allow PRO subjects, as in adult infinitives. Surprisingly however, in bare stems (which are also missing T), overt subjects occur at approximately 12%. I provide an analysis that makes use of a null constant-anaphoric topic operator construction (Rizzi, 1992; 1997). I show that in both adult and child Swahili, this construction occurs in the absence of agreement. It is this anaphoric topic operator (and not a true subject) that occurs in both [-SA] clauses (adult and child) as well as child bare stems.

1.1 The study

The focus of this study is the acquisition of verbal morphology in Swahili, an Eastern Bantu language with heavy agglutinative prefixing and suffixing. This is the first investigation of the acquisition of Swahili, and indeed the first study of the acquisition of any eastern Bantu language. While a handful of other Bantu languages have been studied, those results differ significantly from what is reported here. Similarly, relatively few agglutinative languages have been studied, and so this study of Swahili contributes to this neglected area of research. Finally, most of what we know about the acquisition of affixes is about the acquisition of suffixes. We have a general idea that the acquisition of prefixes is problematic for children (e.g., Slobin, 1973), but very little empirical data is available. This study aims to contribute to our knowledge of the acquisition of prefixes.

In this chapter I will first discuss the fact that children invariably converge on the morphosyntactic system of their language early in the developmental sequence. Hoekstra & Hyams (1998) discuss this as Early Morphosyntactic Convergence (EMC), and Wexler (1998) describes this as Very Early Knowledge of Inflection (VEKI). I will suggest that child language falls into three classes based on how verbal inflection is used: Root Infinitive (RI) languages (such as German and Dutch, see below for a description of what an RI is), non-Root Infinitive languages (such as Italian

and Spanish), and bare verb languages (such as English). This is summarized in table 1.1 below¹:

Table 1.1 Summary of languages that allow RIs, disallow RIs and allow bare verbs.

True RI languages	Non-RI languages	Bare Verb Languages
German	Italian	English
Dutch	Spanish	
French	Catalan	
	Japanese	

Hoekstra & Hyams (1998) show that there are morphosyntactic and semantic differences between RIs and bare verbs. The only bare verb language that they consider is English, but I will show that bare verb languages are considerably more frequent than commonly thought. I will outline some previous findings from other Bantu languages, focusing on the acquisition of verbal morphology. I will show that Bantu languages fall into the category of bare verb languages. I will then discuss the acquisition of two agglutinative languages, and we will see that the omission of verbal morphology is pervasive in these languages too. This typology will serve as background for our investigation of the acquisition of Swahili.

We will see that Swahili children allow bare verbs like the children of other Bantu languages. However, because Tense in Swahili is obligatory (in contrast to languages such as Sesotho and Siswati), it allows us to investigate the development of several inflectional affixes at the same time. This in fact will be the most revealing aspect of this study – that subject

¹ In addition there are languages that allow bare participles, default finite forms, etc. I ignore these here for the sake of simplicity.

agreement and tense develop differentially and have different effects on other aspects of grammar, e.g., the occurrence of overt subjects. We will see that Swahili children allow bare verbs, but also allow various other sorts of underspecified verbs. I show that these underspecifications are accounted for by a model of underspecification called ATOM (Agr-Tense Omission Model, Schütze & Wexler, 1996).

Let us begin with an overview of the types of errors children make. I will first show that errors of omission are very frequent and typical of child language. One product of such errors is the bare verb in English. I will then show that errors of commission are relatively rare in child language, with the apparent exception of Root Infinitives in languages such as German². RIs have received so much attention in the literature, and so I will describe their characteristics in section 1.2.3. With the addition of languages that allow neither bare verbs nor root infinitives (e.g., Italian), this will establish our typology in table 1.1 above.

1.2 Early Morphosyntactic Convergence

Children don't begin talking until around the first birthday, after which it takes them several years to fully acquire the grammar of their ambient language. While it may take several years for them to gain mastery of their language, during that time children's grammar does not develop at random. Children make systematic errors and progress through systematic stages in the acquisition of language, culminating in mastery of the language they are born into. One such stage in language acquisition that

begins with multi-word utterances and ends sometime around the 3rd birthday is the so-called Telegraphic Stage (Brown, 1973). Telegraphic speech is characterized by the omission of obligatory elements such as determiners, prepositions, agreement markers, etc.

1.2.1 The Frequency of Omission

Children acquiring various languages omit determiners (1), copulas (2), auxiliaries (3), subject-verb agreement (4, taken from CHILDES, MacWhinney (2000), etc.:

- | | | |
|-----|---|--|
| (1) | Determiner Omission | |
| a. | Paula play ball | Paula, 1;6 (Radford, 1990) |
| b. | Haley draw boat | Hayley, 1;8 (Radford, 1990) |
| c. | want open door | Daniel, 1;8 (Radford, 1990) |
| d. | Niekje ook boot maken
Niekje also boat make
'Niekje has also made a boat' | Dutch (Schaeffer, 1994) |
| e. | Papa heeft ook trein
Daddy has also train
'Daddy also has a train' | Dutch (Schaeffer, 1994) |
| f. | est tombé éfant
is fallen elephant
'The elephant has fallen | French (Ferdinand, 1996) |
| g. | train va tomb[e]
train go fall
'The train is going to fall.' | French (Ferdinand, 1996) |
| (2) | Copula Omission | |
| a. | I in the kitchen | English (Becker, 2000) |
| b. | Da rote ball
there red ball | German (Salustri & Berger-Morales, 2001) |

² I say *apparent*, because I take the position (as many in the field) that RIs and bare verbs arise from the same underlying mechanism, namely the underspecification of functional heads. See chapter 4 for details.

- (3) Auxiliary Omission
- a. baby talking Hayley, 1;8 (Radford, 1990)
 - b. doggy barking Bethan, 1;9 (Radford, 1990)
 - c. Mummy doing dinner Daniel, 1;10 (Radford, 1990)
- (4) Subject-Verb Agreement Omission
- a. It only write on the pad Eve, 2;0 (Brown, 1973)
 - b. Cromer have some Adam, 2;7 (Brown, 1973)
 - c. He bite me Sarah, 2;9 (Brown, 1973)

Note that the examples in (4) exemplify the fact that English children allow bare verbs in contexts in which an agreement marker is obligatory (3rd person singular –s). Not only is omission typical, it is very frequent overall. Table 1.2 below shows that the omission of 3rd person singular agreement in English occurs at rates of approximately 80%:

Table 1.2 Proportion of bare verbs in child English

Child	Age	% Bare Verbs
Eve	1;6-1;10	78
Adam	2;3-3;0	81
Nina	2;4-2;5	75

(Sano & Hyams, 1994)

We will see shortly that in a variety of agglutinative languages too, omission is extremely frequent.

1.2.2 The Paucity of Agreement Errors

While omission is frequent and widespread, errors of agreement are extremely rare. Table 1.3 (from Sano & Hyams, 1994) shows that children in a variety of languages generally make errors of agreement less than 4% of the time.

Table 1.3 Agreement errors in various languages

Child	Language	Age	n	%error	Source
Simone	German	1;7-2;8	1732	1%	Clahsen & Penke 1992
Martina*	Italian	1;8-2;7	478	1.6%	Guasti 1994
Diana*	Italian	1;10-2;6	610	1.5%	Guasti 1994
Guglielmo*	Italian	2;2-2;7	201	3.3%	Guasti 1994
Claudia	Italian	1;4-2;4	1410	3%	Pizzuto & Caselli 1992
Francesco	Italian	1;5-2;10	1264	2%	Pizzuto & Caselli 1992
Marco	Italian	1;5-3;0	415	4%	Pizzuto & Caselli 1992
Marti*	Cat/Spanish	1;9-2;5	178	0.56%	Torrens 1992
Josep*	Cat/Spanish	1;9-2;6	136	3%	Torrens 1992
Gisela*	Catalan	1;10-2;6	81	1.2%	Torrens 1992
Guillem*	Catalan	1;9-2;6	129	2.3%	Torrens 1992

(*data available on CHILDES, MacWhinney, 2000: Martin, Guglielmo, Diana corpora, Cipriani *et al.* 1991; Marti, Josep, Guillem, Gisela corpora: Serra & Solé, 1992)

Take English for example, where children hardly ever use 3rd person singular agreement morphology in 1st person singular contexts, as in the hypothetical example given below:

- (5) I eats cake UNATTESTED

Table 1.4 below shows that from the data of 10 children, Harris & Wexler (1996) identified 1724 verbs that occur in 1st person singular context, of which only 3 occur with the incorrect 3rd person singular –s suffix.

Table 1.4 Frequency of verbs in first person singular contexts

Stem	Irregular Past	-ed	-s
1349	325	47	3

Data from 10 children on CHILDES, age range = 1;6-4;1 (Harris & Wexler, 1996)

Similarly, Schieffelin (1985) shows that children acquiring Kaluli ‘very rarely put the wrong case marking on nouns’ (p.537). Smoczy_ska (1985) reports the same for Polish speaking children, and Clancy (1985) says that ‘the system of verbal inflections in Japanese emerges quite early, and errors are not frequently reported’ (p.383). Thus such errors of commission are extremely rare, while errors of omission are common.

This general paucity of such errors has led researchers to conclude that children acquire knowledge of the morpho-syntactic properties of their language extremely early, hence Hoekstra & Hyams' (1998) EMC. In the next section we will see a phenomenon that at first glance is an apparent exception to EMC, but which in fact conforms to EMC.

1.2.3 Root Infinitives

The evidence for EMC is that when inflection occurs, it occurs overwhelmingly correctly. However, the occurrence of Root Infinitives casts doubt on the validity of this generalization. A Root Infinitive (RI) is a verb that occurs with infinitival morphology in a matrix clause (Hoekstra & Hyams, 1998). Thus, instead of having finite inflectional morphology in a main clause, RIs have 'incorrect' infinitival morphology, as in the examples in (6)³. In example (6a), the child uses the infinitive form of the verb *haben* 'to have'. The appropriate finite form of this verb is *hat*. This is not an error of *omission*, but appears to be an error of *commission*⁴. Thus, this may represent an exception to EMC.

³ This view that RIs are verbs with infinitival morphology is made explicit by Hoekstra & Hyams (1998). However, others propose that RIs occur as default forms when the verb is underspecified for Tense (Wexler, 1994), Agr (Clahsen, Eisenbeiss & Penke, 1996), or both Agr and Tense (Schütze, 1997). See chapter 4 for a discussion of these proposals.

⁴ Carson Schütze points out that 3rd singular subject is not incompatible with the infinitive, and thus is not an error of commission in that sense. However, it is an error of commission in the sense that there is a mismatch between the morphology (non-finite) and the intended meaning (finite). The English bare form does not fit this criterion of being an error of commission because while the form does not match the intended meaning, there is a total absence of morphology, thus an error of omission.

- | | | |
|---|---|--------|
| (6)a. Thorstn das haben
Thorstn that have-inf
'Thorstn has that.' | b. Zahne putzen
teeth brush-inf.
'(Someone) brushes (his)
teeth.' | German |
| c. Papa schoen wassen
daddy shoes wash-inf.
'Daddy washes (the) shoes.' | d. Ik ook lezen
I also read-inf.
'I also read.' | Dutch |
| e. Fermer yeux
close-inf. Eyes
'(I have) closed (my) eyes' | f. 'Tasha ouvrir
Natasha open-inf.
'Natasha is opening/
going to open/wants to open it.' | French |

RIs occur in languages such as German (Weissenborn, 1990; Poeppel & Wexler, 1993), Dutch (Weverink, 1989), French (Pierce, 1989), Swedish (Platzack, 1990), etc.

Table 1.5 Frequencies of RIs in French and Swedish

Language	Child	Age	%RIs
French	Nathalie	1;7-2;11	76%
French	Daniel	1;5-2;5	60%
Swedish	Freja	1;11-2;0	38%
Swedish	Tor	1;11-2;2	56%
Swedish	Embla	1;8-1;10	61%

(Adapted from Sano & Hyams, 1994)

It is important for us to discuss the various characteristics of RIs because they are such a pervasive phenomenon in several languages in Western Europe. Because RIs are attested in some child languages, a priori we expect RIs to be a possibility for Swahili children too. RIs have become an important focus of study in the acquisition of European languages, precisely because they represent an apparent departure from the EMC. However, upon closer inspection, it becomes clear that they are not an exception to the EMC. RIs have several defining characteristics. First, RIs

are optional: RIs occur side-by-side with fully inflected verbs, sometimes in immediately adjacent utterances. Examples (7-8) are taken from a single corpus of a German speaking child, Andreas (cited in Wexler, 1994)⁵.

- (7) [+finite] verbs
- a. Mein Hubsabe hat Tiere din
my helicopter has animals in it
'My helicopter has animals in it'
- b. Caesar tieg e nich ('Caesar kriegt er nicht')
Caesar get he not
'He is not getting Caesar.'
- (8) [-finite] verbs
- a. ich der Fos hab'n
I the frog have (-fin)
'I have the frog.'
- b. Thorstn das haben
Thorstn that have (-fin)
'Thrstn has that.'
- c. Zahne pussen
teeth brush (-fin)
'I (want) to brush my teeth.'
- d. tein ('kein') Zahnburste liegen
no toothbrush lie (-fin)

Behrens (1993) investigates temporal marking in the speech of Simone (Miller, 1976), a child acquiring German, and reports that RIs occur

⁵ The translations here are those provided by the authors. They represent the authors 'hypothesis as to what the child intended. We will see shortly that others have proposed that RIs generally have modal meanings associated with them. The translations given in the text would thus change so as to incorporate a modal meaning.

alongside finite verbs in the same transcripts. Her figures are given in table 1.6.

Table 1.6 RIs and finite verbs in the files of Simone

Age	RIs	Finite
2;0	72%	28%
2;6	32%	68%
3;0	16%	84%
4;0	4%	96%

(Behrens, 1993)

Second, it is well known that RIs occur in the same position as infinitival verbs in the adult language, namely in an *unraised* position. In child German, RIs occur overwhelmingly in final position, while finite verbs in child German occur overwhelmingly in second position (examples and table taken from Poeppel & Wexler, 1993):

- (9) a. Thorsten Caesar haben unraised RI (SOV)
Thorsten Caesar have-inf
'Thorsten has (the doll) Caesar'
- b. Ich hab tein Bürse raised finite verb (SVO)
I have (a) small brush

Table 1.7 Finiteness versus verb position in German

	+finite	-finite
V2	197	6
V final	11	37

(Table adapted from Poeppel & Wexler, 1993, p.7. Data is from Andreas, age 2;1, taken from the CHILDES system, MacWhinney, 2000)

Similarly, adult French has verb raising over negation and adverbs when the verb is finite, but not when the verb is infinitival. Pierce (1989) reports that French children produce RIs in the unraised position (post negation), and inflected finite verbs in raised position.

- (10) a. pas casser unraised RI (V-neg)
not break
- b. marche pas raised finite verb (neg-V)
walks not

Table 1.8 Finiteness versus verb position in French

	+finite	-finite
Verb <i>pas</i>	216	2
<i>pas</i> Verb	9	122

(Table is from Pierce, 1989; data are from four French speaking children aged 1;8 – 2;2)

This correlation between verb raising and RIs strongly refutes the view that RIs are speech errors. Because child RIs behave syntactically like infinitives in the adult language, this means that RIs are syntactic infinitives that are permitted in root clauses and they give rise to certain interpretations. Adult languages too allow RIs with particular interpretations (see Schütze, 1997; Hoekstra & Hyams, 1998 for details), but not in declarative contexts that we see RIs in child language. More importantly, however, the fact that RIs behave like adult infinitives shows that they are not exceptions to the EMC. If children have not converged on the morphological properties of infinitives (and the associated syntactic restrictions on that morphology), then they should not know the position requirements of infinitives. As the two tables above show, children clearly have knowledge that RIs are infinitives.

The third characteristic of RIs is that they generally occur with null subjects (Weverink, 1989; Pierce, 1989; Phillips, 1995).

Table 1.9 Null subjects in finite and non-finite clauses in several child languages

		Null Subjects in Finite Clauses	Null subjects in non-Finite clauses
French (Pierce, 1989)	Daniel	150/273 (55%)	166/205 (81%)
	Nathalie	90/304 (30%)	131/295 (44%)
	Philippe	182/782 (23%)	153/194 (79%)
German (Behrens, 1993; Krämer, 1993)	Simone	781/3699 (21%)	2199/2477 (89%)
	Andreas	34/263 (13%)	69/101 (68%)
Dutch (Krämer, 1993; Haegeman, 1995)	Thomas	165/596 (28%)	246/267 (92%)
	Heinz	1199/3768 (32%)	615/721 (85%)
Flemish (Krämer, 1993)	Maarten	23/92 (25%)	89/100 (89%)
Faroese (Jonas, 1995)	O.	8/52 (15%)	67/161 (42%)
Danish (Hamann & Plunkett, 1998)	Anne	366/3379 (11%)	394/667 (59%)
	Jens	742/3173 (23%)	539/937 (58%)

(Adapted from Rasetti, 2000, table 9.1, p.239).

In table 1.9 we see that the proportion of null subjects in non-finite clauses is consistently higher than the proportion of null subjects in finite clauses in a range of languages. This is reminiscent of the correlation between adult non-finite embedded clauses and null PRO subjects, and thus it has been suggested that child RI clauses are like adult infinitival clauses (e.g., Guilfoyle, 1984; Sano & Hyams, 1994; Radford, 1990; Hyams, 1996). See chapter 5 for more discussion of null subjects and their relation to underspecified clauses. Because child RIs have this property of adult infinitives, this is further evidence that RIs are not an exception to the EMC.

The fourth characteristic of RIs is that they generally occur in modal contexts (for Dutch, see Wijnen, 1996; for French see Ferdinand, 1996; for Swedish see Plunkett & Strömquist, 1990; for a review see

Hoekstra & Hyams, 1998). Examples are given below (from Wijnen, 1996):

- (11) a. Papa boek lezen
Daddy book read-inf
'(I want)Daddy to read the book'
- b. Niekje buiten spelen
Niekje outside play-inf
'Niekje wants to play outside'

Hoekstra & Hyams (1998) argue that this modality comes from the presence of infinitive morphology, which cross-linguistically is generally associated with modal meaning (Duffley, 1992; Bolinger, 1968; Stowell, 1981; Han, 2000). The fact that child RIs are associated with modal/irrealis contexts is further evidence that the infinitival morphology on RIs is not 'incorrect' use of morphology, but rather that RIs are genuine syntactic infinitives, and thus do not represent an exception to the EMC.⁶

⁶ The fifth characteristic of RIs is that they generally do not occur in wh-contexts. While they are very frequent in declarative clauses, it has been shown for several languages that RIs are very rare in wh-contexts (e.g., Kursawe (1994) for German; Haegeman (1994) for Dutch). The table below shows that in the corpus of Hein (Elbers & Wijnen, 1992), of the 721 non-finite verbal clauses, only 2 occur with wh-questions (0.2%). However, wh-questions occur at a rate of 2.3% in finite contexts. This does not directly refute the view that RIs are exceptions to EMC, but is an important characteristic that we will return to in chapter 5.

Finiteness in declaratives versus wh- questions in child Dutch

	+finite	-finite
All Clauses	3768	721
Wh- questions	88	2

Table taken from Haegeman (1994); data is from Hein (Elbers & Wijnen, 1992)

However, not all languages allow RIs. Sano & Hyams (1994) show that RIs are very frequent in languages such as German, Dutch, French, Swedish, etc., but are somewhat rarer in languages such as Italian, Spanish, Catalan, etc. (see table 1.10 below). Children acquiring these languages produce neither RIs nor bare verbs (see for example Guasti, 1992). Thus these child languages are classified as a third group, distinct from RI languages and bare verb languages.

Table 1.10 Proportion of RIs in some non-RI languages

Language	Child	Age	%RI
Italian (Guasti, 1992)	Diana	2;0	0%
	Martina	1;11	16%
		2;1	4%
Italian (Schaeffer, 1990)	Paola	2;0-2;5	7%
	Daniele	1;7-2;6	8%
	Massimo	1;7-2;6	6%
	Gabriele	1;7-2;6	7%
	Orietta	1;7-2;6	5%
	Elisabeth	1;7-2;5	10%
Spanish (Grinstead, 1994)	Damariz	2;6-2;8	5%
	Juan	1;7-2;0	12%
		2;1-2;4	10%
Catalan (Torrens, 1992)	Guillem	1;11-2;6	3%
	Marti	2;0-2;5	3%

(Adapted from Sano & Hyams, 1994)

Furthermore, there is the issue of whether English bare verbs are RIs, or whether they should be categorized separately. Wexler (1994) suggests that English bare verbs are the result of the same processes that produce German and Dutch RIs, namely the underspecification of tense. Wexler argues that one appealing reason to assimilate them to RIs is that it

would unify English (which at that time was the only bare verb language that was well known) with other Germanic languages. However, we will see that bare verbs occur frequently in the languages of the world, and thus categorizing English bare verbs as RIs obscures this distinction. While bare verbs and RIs may be the result of the same underlying mechanism, it is unclear why in some languages bare verbs are attested while in other languages RIs are attested (and still in other languages neither are attested).

I will not provide an answer to this question, but there are a number of possibilities that come to mind that may be worth pursuing in the future. One possibility is the morphological differences between Bare verb languages and RI languages. Specifically, it has been noted that infinitives in languages such as German and Dutch behave differently and are of a different morphological status than the infinitive in English. For example, in the adult RI languages the infinitive occurs as a postverbal affix while in English it occurs preverbally. Furthermore, as Schütze (1997) points out, the infinitive in RI languages occurs as a bound morpheme to the verb, with no intervening lexical material allowed. However, in English the infinitive marker has a somewhat looser relationship with the verb, since intervening lexical material is allowed, for example “To boldly go where no one has gone before”.

However, this does not answer the question of why languages such as Italian and Spanish (which have bound postverbal infinitives) do not exhibit RIs. Thus the reason for this three-way typology remains a mystery. What I will show in the next section is that Bantu languages disallow RIs and allow bare verbs, and thus are categorized as Bare Verb Languages. The infinitive in Bantu languages resembles the English infinitive in that it is preverbal and allows intervening material, suggesting that the preverbal

position and free nature of the infinitive perhaps is related to the occurrence of bare verbs.

1.3 Acquisition of Bantu Languages

In this section I will briefly discuss some previous acquisition findings from two other Bantu languages: Sesotho and Siswati. The only acquisition studies reported in the literature are of southern Bantu languages, such as the pioneering work on Sesotho by Katherine Demuth. The majority of what is discussed in this section comes from work done by Demuth and her colleagues. For an excellent review of the acquisition of Sesotho and the Bantu acquisition literature, see Demuth (1992b).

1.3.1 Sesotho

Sesotho is a southern Bantu language spoken in the nation of Lesotho. Like Swahili it has rich noun class morphology, but unlike Swahili, Sesotho is a tonal language. Sesotho also has a rich verbal complex, with the word order given in (12) below:

(12) Subject SA–(T/A)–(OA)–Verb–suffixes/perf. /pass.–Mood Object

The verbal complex consists of a subject agreement marker, followed by an optional tense/aspect marker, an optional object agreement marker, various suffixes (including grammatical function changing suffixes), an optional aspect marker, and a mood final vowel. An example is given below (taken from Demuth, 1992)⁷:

⁷ I follow Demuth’s transcription protocol, where ´ indicates high tone, and low tone is left unmarked.

- (13) Thabo ó – pheh – íl – é di–jó
 Thabo SA₁–cook–PERF–MOOD 8-food
 ‘Thabo cooked (some/the) food.’

Sesotho is a tonal language, making use of two tones (Demuth, 1993): high, and Ø (which surfaces as a default low). Unlike Swahili, the only prefix that must occur in every declarative utterance is the subject agreement prefix. As we will see in chapter 2, Swahili requires subject agreement as well as tense in almost every declarative context.

The infinitive in Sesotho (and indeed in Swahili as well) is preverbal and may occur with object agreement intervening between it and the verb. Thus morphologically it resembles the English infinitive in these ways. However, lexical material may not intervene between the infinitive prefix and the verb, suggesting a somewhat tighter relationship than in English.

The inflectional prefixes are the focus of the current study, and so I will restrict the discussion to the acquisition of verbal prefixes in Sesotho. Demuth reports that children acquiring Sesotho go through three non-discrete stages. In the first stage, children predominantly produce bare verbs, that is, verbs that are missing all inflectional prefixes. The second stage is typified by verbs which have ‘shadow vowel’ prefixes. Such prefixes are typically [a] or [e] in Sesotho. This phenomenon has been noted in other Bantu languages as well, notably Siswati (Kunene, 1979, see below). (14) is an example of an utterance with a shadow vowel from Sesotho. The first line is the child utterance, followed by the adult equivalent, a gloss and a translation.

- (14) a lahlíle
 ke – di – láhl – íl – e+
 SA_{1s} –OA₁₀–throw away–PERF–MOOD
 ‘I threw them away.’

In fact, such preverbal syllables are not uncommon in the acquisition of other languages. In the literature, these preverbal vowels have been variously labeled: Peters (2001) refers to them as filler syllables; Bottari, Cipriani & Chilosi (1993) refer to them as Monosyllabic Place Holders; Veneziano & Sinclair (2000) call them ‘additional elements’. We will return to this phenomenon in chapter 3 when we investigate child Swahili.

In the third stage, fully inflected forms predominate. Thus the developmental sequence is given in (15).

- (15) no inflection > ‘shadow vowels’ > well-formed inflection

Demuth notes that these stages are not discrete, and that utterances of all three kinds occur in all three stages. The examples below (Katherine Demuth, p.c.) all come from the same child (L, age 2;1), and are attempts at the same utterance. Examples (16a-d) are the child utterances, and example (17) is the adult target.

- (16) a. qetile
 b. ketile
 c. eketile
 d. aketile

- (17) ke – qet – íl – e
 SA–finish–prf–IND
 ‘I finished’

(16a,b) are examples of bare verb stems (i.e., missing prefixes, although they do contain the aspectual suffix), while (16c,d) are examples of utterances with shadow vowels. At the same time, this child also has well-formed subject agreement markers in other utterances:

- (18) ke – i – thol – ets – e tsena
 SA–Rflex–find–APPL/PRF–IND these
 ‘I found these’

However, importantly, while bare verbs do occur in Sesotho, RIs do not. Thus we can classify Sesotho in our typology of child languages as a Bare Verb language, since there is a stage at which children allow bare verbs. These bare verbs occur alongside fully inflected verbs, and are thus typical errors of omission. Demuth reports that bare verbs are significantly less frequent than bare nouns in her corpora, but still do occur. She predicts that bare verbs are more common with younger children.⁸

1.3.2 Siswati

Kunene (1979) investigates the acquisition of noun class morphology, possessive morphology and agreement in Siswati, a southern Bantu language closely related to Sesotho. Her data comes from two children aged 2;2-3;0 and 2;11-3;6. She finds that Siswati children at early

⁸ Besides the acquisition of prefixes, Demuth reports on the acquisition of other verbal elements. For example, Demuth (1989) reports that passives occur relatively frequently in Sesotho speaking children's everyday speech. She finds that, contrary to English speaking children (cf. Borer & Wexler, 1987; Fox & Grodzinsky, 1998), Sesotho children show evidence of spontaneous production of passives as early as 2;8, as well as comprehension of passives. This is different from children acquiring Western European languages, who acquire the passive relatively late. Demuth attributes this early proficiency to the unusual frequency of passives in the input. Sesotho care-givers use an unusually large proportion of passives when speaking to children (in Demuth 1992b the proportion of passives in the care-givers utterances is reported as 23/386 (6%)). Demuth also reports on the acquisition of grammatical function changing suffixes such as the applicative (Demuth, 1998) and the tonal system of Sesotho (1993), as well as nominal morphology (Demuth, 1988, 1992, 1994). However, since the focus of the present study is the verbal complex in Swahili, I will not discuss these results. For an overview of these results see Demuth (1992b).

stages produce bare verbs exclusively. Adult Siswati requires minimally that the verb occur with a subject agreement prefix (tense and object agreement are not obligatory). The youngest child at her first data point produced only bare verbs with no inflectional prefixes whatsoever.⁹ Examples are given below.¹⁰

- (19) a. Zanele lala
 adult form: Zanele u-ya-lala
 Zanele SA-ya-sleep
 'Zanele is sleeping.'
- b. landzela mine
 adult form: ngi - ta -ku -landzela mine
 SA-future-OA-follow me
 'I will follow you.'
- c. tfwana khala
 adult form: um-tfwana u - ya - khala
 pref-child SA-ya-cry
 'The baby is crying.'

These examples are taken from the first recording of the youngest child. In the second recording, when the youngest child was 2;3, she began to produce verbs that had a "rudimentary" subject agreement marker. This "rudimentary" prefix is similar to what Demuth describes in the

⁹ Kunene does not provide quantitative data, so we do not know how many verbal utterances the child produced at this stage.

¹⁰ In two of these examples, there is a prefix 'ya' in the adult form, which Kunene describes as occurring 'when there is no adjunct following the verb.' The exact function of this prefix is unclear, since in the examples she glosses this prefix as 'ya'. I assume it is a marker of intransitivity.

development of Sesotho as a “shadow vowel”. In Siswati, these prefixes occur primarily as [i] or [a], as examples below show:

- (20) a. i/a – lhala la mlumbi
 Adult form: u – lhala la um – lumbi
 SA – sit here pref. – whiteman
 ‘The whiteman sits here.’
- b. a – buka tfombe t – a – mi
 ngi–buka ti–tfombe t – a – mi
 SA – look pref.–picture PA–PM–my¹¹
 ‘I am looking at my pictures.’

In example (20a), ‘i/a’ indicates that the child used both prefixes in free variation. In (20b), only [a] occurred as the “rudimentary” prefix.

Kunene describes the development of prefixes on the verb in Siswati as being roughly parallel to what Demuth describes for Sesotho, i.e., beginning with no inflection, then developing into a shadow/rudimentary vowel, followed by full well-formedness. We will see that early Swahili is similar to Sesotho/Siswati in that it allows ‘shadow vowels’ as prefixes at a particular stage in development. These ‘shadow vowels’ (in Swahili as well as Sesotho and Siswati) are not well-formed in that they do not conform phonologically to any target prefix. The same holds true in Swahili.

Unlike Sesotho and Siswati, Swahili obligatorily takes subject agreement and tense prefixes. Thus, the use of a single (‘shadow’) preverbal vowel (21a) can be contrasted with the occurrence of both well-formed target prefixes (21b), the occurrence of a single well-formed prefix (21c-d), and the occurrence of no well-formed prefixes (21e).

¹¹ PA = possessive agreement marker, PM = possessive morpheme

- (21) Possible prefixes in Swahili:
- a. preverbal shadow vowel
 - b. Both SA and T prefixes
 - c. Only SA prefix
 - d. Only T prefix
 - e. Neither SA nor T prefixes

This is of relevance in the current study because we will see that Swahili children in fact produce each of these possibilities.

The Bantu languages that have been studied to date are all southern Bantu languages. I have only reported on Sesotho and Siswati, but other languages that have been studied include Zulu (Suzman, 1982; 1991), Basotho (Connelly, 1984; 1987), and Chichewa (Chimombo, 1981).¹² These languages differ from Swahili in that tense is not obligatory: in Swahili, no declarative clause may occur without a tense prefix. The fact that Subject Agreement and Tense are marked as independent prefixes in Swahili in almost every declarative clause sets it apart from the other Bantu languages mentioned earlier in this chapter. The acquisition of these two prefixes will be the primary theme of this dissertation. I turn now to a brief look at the acquisition of other agglutinative languages.

1.4 Other Agglutinative Languages

I will focus on the acquisition of verbal affixes, and we will see that children generally allow omission of these affixes. We will see that children identify the bare verb stem at early ages, and morphology is added

¹² Chichewa is not from the same group as Sesotho or Siswati. However, the study cited focuses on negation, and the differences between first and second language acquisition. Other work by Chimombo has focused on tone, and thus are not of direct relevance to Swahili.

to that stem as children develop. This is the case in Sesotho and Siswati, and we will see this in Quechua and Navajo. This will establish a pattern in child language that allows bare verb stems.

1.4.1 Quechua

Courtney (1998) investigates the acquisition of Quechua in the spontaneous speech of four children aged 2;0-2;8. Quechua is an SOV agglutinative language with a verbal complex that has extensive suffixing. In adult Quechua, the order of morphemes is given in (22)¹³:

(22) V – Derivational – OA – Prog – OA – T – SA – Num – Conditional

The derivational suffixes include transformative, factitive, desiderative, and perdurative. OA occurs either after the derivational suffixes or after the progressive marker, and the SA marker is compositional between person and number marking. Furthermore, “a Quechua verb must bear either a person-of-subject suffix appended to the root, or the Imperative/Infinitive morpheme. That is to say, adult Quechua speakers never produce bare roots or stems.” (Courtney, 1998, p.60). For an imperative, the infinitive/imperative suffix is appended at the end. Thus the infinitive occurs as a suffix in Quechua (unlike either English or Bantu languages). As can be seen from (22) several suffixes may intervene between the verb and the infinitive suffix.

In Quechua child-directed speech, Courtney reports that of 279 identifiable verbs, 249 occur with SA as well as at least one additional suffix (p.168). Therefore almost 90% of all verbs that the children hear

¹³ Derivational = derivational suffix; OA=Object Agreement; Prog = progressive aspect ; T = Tense ; SA = Subject Agreement ; Num = number marking

occur with at least two suffixes. The remaining 30 verbs occur with SA. Importantly, no bare verbs occur in the input.

Children, on the other hand, allow bare verbs quite frequently. The youngest child (2;0-2;2) produced 81 verbs (with 48 different verb roots), of which 46 were bare.¹⁴ The second child (2;5-2;7) produced 37% bare verbs in the first half of his corpus, but in the second half produced none whatsoever. The third child (2;7-2;8) also produced no bare verbs. Thus bare verbs are a symptom of very early child language (as Demuth surmises about child Sesotho). Examples are given in (23). In each case, the child produces the bare verb (*muna*) when the adult form would have taken at least one suffix.¹⁵

- (23)
- | | |
|-------------|---|
| a. | Chay muna qan |
| adult form: | chayta munanki qan
'you want that' |
| b. | Carruta muna noqa |
| adult form: | noqa carruta munani
'I want the car' |
| c. | noqa carru muna |
| adult form: | noqa carruta munani
'I want the car' |
| d. | mana muna noqa carrupi |
| adult form: | noqa carrupi mana munanichu
'I don't want the car' |

¹⁴ The results are as follows: 46 bare verbs (57%), 20 V+SA (25%), 19 V+ Infinitive/Imperative (23%), with the remaining 14 being of all different sorts.

¹⁵ Note that the subject and object also differ from the adult form in that case marking is often omitted. This is relevant only insofar as it further exemplifies the tendency towards omission. However, our focus here is on the acquisition of verbal morphology.

Recall that Quechua does have an infinitive marker that occurs at the end of the verbal complex. RIs, however, are not attested. 19 of the youngest child's 81 verbs occurred with the infinitive/imperative marker on it, but they were used in imperative contexts. Therefore these 19 utterances were used correctly, and not as unadult-like root clause infinitives.

1.4.2 Inuktitut

Inuktitut is a polysynthetic language spoken in arctic Quebec. The adult language does not permit bare verbs, and does not have a morphological infinitive. Swift & Allen (2002) report that normally developing Inuit children allow bare verbs that are ungrammatical in the target language. The typical verbal complex is made up of the verb root, several optional suffixes (e.g., modalization, negation, etc.), followed by an obligatory inflectional suffix. This final inflectional suffix is a portmanteau morpheme encoding person, number and mood.

Swift & Allen take their data from four Inuit children aged between 2;0 and 3;6. They found that while verbal inflection dropping occurs in adult speech, it is far more frequent in child speech than adult speech:

Table 1.11 Verbal Inflection dropping by Inuit children and adults

	Children	Caretakers
Uninflected verbs	145 (5.6%)	7 (0.29%)
Inflected verbs	2439	2409

They found that overall the omission of verbal inflection decreased with maturity, and that often the children produced verbs in both inflected and uninflected forms. The following is from the same child, aged 2;10:

- (24) a. piilaurit
 piiq – lauq – git
 remove–POL.IMP–2sS
 ‘You get off, please’
- b. pii!
 Piiq – Ø
 Remove–no.infl.
 ‘(you) get off.’

Swift & Allen conclude that the omission of verbal inflection cannot be due to a lack of knowledge of the inflection itself, but must be due to other factors. They investigate whether omission of inflection is due to discourse-pragmatic conditions, structural conditions and/or emotional conditions. They find that none of these conditions alone account for all verbal omission, suggesting that inflection omission is generally a much more complex phenomenon than is commonly assumed.

The details of their study are not crucial for our purposes here, simply the fact that Inuit children omit verbal inflection and allow bare

verbs at significantly higher rates than adult Inuktitut allows¹⁶. Thus Inuktitut is another language that can be added to the bare verb category.¹⁷

Thus we can add these languages to the inventory of languages that we saw in table 1.1 earlier.

Table 1.12 Summary of languages that allow RIs, disallow RIs and allow bare verbs.

True RI languages	Non-RI languages	Bare Verb Languages
German	Italian	English
Dutch	Spanish	Sesotho
French	Catalan	Siswati
Swedish	Japanese	Quechua
Icelandic		Inuktitut
Russian		

Our conclusion from this discussion then is that bare verbs are not restricted to English alone. English is thus one member of a class of diverse

¹⁶ Adult Inuktitut arguably does not allow bare verbs at all, given the rate of 0.29% cited by Swift & Allen. However, it is unclear from Swift & Allen's study whether adult native speakers disallow bare verbs in all discourse contexts.

¹⁷ Feurer (1980) describes the acquisition of Mohawk verbal morphology. The development of Mohawk verbal morphology is described as 'an expansion' from bare, minimal verbs to fully adult-like verbs. This, as we have seen in all the languages discussed so far, is the common pattern. In Mohawk, as in Navajo, the bare stem is unattested in the adult language. There is minimally one obligatory prefix which marks subject agreement. Examples are given of child utterances that are entirely missing any prefix. In addition, Feurer describes utterances that have a proto-prefix, which can be assimilated to Demuth's 'shadow vowel' (although in Mohawk such prefixes are not always vowels). The age of the child being described is 3;3 at the earliest stage, and it is only at this stage that the child produces bare verbs. I do not include a comprehensive review of Mohawk because the data and description in Feurer (1980) is limited.

a.	konó:lu	b.	kyó?teh
target form:	yo-kvno:ru	target form:	wa-kyó?te?
	it-is-raining		I-am-working

languages, and assimilating English to RI languages misses this fact. One fact that this table fails to capture is that Sesotho, Siswati, Quechua, etc. differ from English in that the former group of languages have true infinitive markers (as RI languages). English, on the other hand, arguably does not have a real infinitive marker since 'to' behaves differently from infinitive markers in other languages (see chapter 2 for more discussion of this point, specifically the differences between French infinitives and English 'infinitives. See also Schütze, 1997). We will return to this point in chapter 5 in the conclusion.

The present study is concerned with the acquisition of Swahili, the acquisition of which we know very little about. We will see that Swahili has an infinitive prefix (*ku*), and so RIs in principle should be available to children. However, Swahili may behave like Italian (which also has an infinitive, but which is not an RI language). Swahili may behave like the closely related language Sesotho, which allows bare verbs, but not RIs. The inflectional morphology of Swahili is particularly interesting because, unlike Sesotho, it independently and distinctly marks Agr and T. As we will see, this proves crucial in distinguishing among several theories in language acquisition.

In this study, I will describe the phenomenon of omission in Swahili, and show that in the case of Swahili, omission is primarily a syntactic process. I will show that the omission of verbal prefixes by Swahili children has effects on the occurrence of overt subjects. Specifically, we will see that the omission of tense results in the complete absence of subjects in early Swahili, a fact that follows directly from the theory of PRO in adult languages. This confirms previous accounts of null

subjects in child language that posit that null subjects are PRO (e.g., Hyams & Wexler, 1993; Bromberg & Wexler, 1995). I will also argue that in adult Swahili, the omission of Subject Agreement is possible when there is an anaphoric topic operator binding a null constant in subject position. I show that children even at early ages show evidence of having knowledge of this adult possibility. This provides evidence that children acquiring Swahili have access to universal principles of language since the evidence in the input for this construction is minimal. I will add further support to the general consensus that the knowledge that children exhibit about their language is based (at least in part) on universal principles of language.

1.5 Organization of the Dissertation

This dissertation is organized as follows. In chapter 2 I present a description of Swahili. I describe each inflectional affix in detail, as well as derivational suffixes, nominal morphology and some basic phonological facts. I present my theoretical assumptions, and then move on to two current debates in the Bantu literature that deal first with whether agreement prefixes are agreement or pronominal clitics, and second with whether tense in Swahili is a tense prefix or an auxiliary verb. These are active debates in the literature with no absolute consensus. I contribute to this debate because it is necessary in order to understand the status of the prefixes in this dialect of Swahili. I argue that the agreement prefixes (both subject and object) are both agreement and not pronominal, and that tense is a tense prefix, and not an auxiliary verb. I then go on to show that Swahili allows subject agreement to be omitted, and I present a typology of clauses that allow this. I propose that subject agreement omission occurs in the presence of a null constant bound by a topic operator, and I provide

evidence from embedding and quantified subjects. Thus, in addition to NP-traces, *pro* and PRO, adult Swahili has null constant clauses. The analysis that I propose asserts that adult Swahili exhibits several different null elements, which must be acquired by children. We will see in the later chapters that children indeed do acquire these null elements fairly early on.

In Chapter 3 I discuss the methodology, data collection, the subjects, transcription protocols, and the staging criteria. Chapter 4 deals with the emergence of the inflectional prefixes. I first outline some influential theories that deal with omission of inflectional affixes or RIs. I then present the Swahili data. This is the chapter in which most of the empirical results will be described, quantified and exemplified. I then return to the theories of language acquisition and evaluate them in light of the Swahili data.

Chapter 5 addresses the question of subject and subject agreement omission in child Swahili. I show that subject omission correlates with the various underspecified clause types. This is evidence that the omission of inflectional prefixes is not a phonological or purely morphological process, but a process that is syntactic in nature. Furthermore, the omission of subject agreement in child Swahili adheres to the principles of subject agreement omission in adult Swahili that we discuss in chapter 2, showing that subject agreement omission is not simply an error on the part of the child. I show that children exhibit knowledge of various silent elements very early on. The evidence for these silent categories is very sparse in the input, and thus I conclude that children acquire aspects of their language with little or no overt evidence in the input. This is evidence that children acquire language with the aid of universal principles of language.

Chapter 2. Morphosyntax of Swahili

Introduction

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

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

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

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

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

study. I make no claim that Swahili spoken in Nairobi constitutes a single, unified dialect.¹⁸

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

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

¹⁸ The dialects of the families in this study are not significantly different from each other (see appendix 3A). Where this dialect falls on the continuum is not crucial to our purposes here, but I believe it is closer to the non-standard end than to Kiswahili Sanifu. Thanks to Thomas Hinnebusch for discussion on this point.

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

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

2.1 Basic Facts

Swahili is an agglutinative language, with considerable prefixing and suffixing. The unmarked word order is S-V-O, as shown in example (1)¹⁹ below. In (1), the subject (Juma) occurs preverbally and the object (Mariam) occurs postverbally. The verb is embedded in a verbal complex which consists of subject agreement (*a-*) on the left periphery, followed by tense (*-na-*), object agreement (syllabic *-m-*) and then the verb root itself (*pend-*). The verb is followed by (in this case) one suffix which indicates mood (in this case indicative *-a*). The subject can be optionally absent (shown in example 2), and the person and number features of the subject are recoverable from the rich subject verb agreement. The subject may occur in postverbal position (3), with an obligatory pause and lower intonation (so-called comma intonation). Furthermore, the object may also be dropped (4).

¹⁹ Refer to the list of abbreviations after the table of contents for a guide to the glosses. Subscripts indicate agreement features/class between the SA and the subject, or the OA and the object.

- (1)

Subject	Verbal Complex	Object
Juma	a - na - m - pend - a	Mariam
Juma	SA _{3s} -Pres- OA _{3s} - like - IND	Mariam
'Juma likes Mariam'		
- (2) A - na - m - pend - a Mariam
SA_{3s}- Pres- OA_{3s}- like- IND Mariam
'He likes Mariam'
- (3) ni - na - m - pend - a Mariam, mimi
SA_{1s}- Pres- OA_{3s}- like - IND Mariam Spro_{1s}
'I like Mariam'
- (4) a - na - m - pend - a
SA_{3s}-Pres - OA_{3s} - like - IND
'He likes her'

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

- (5) Mariam, Juma a - na - m - pend - a
Mariam, Juma SA_{3s}- Pres- OA_{3s} - like - IND
'Mariam, Juma likes [t]'

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

2.2 General phonological characteristics

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

Most words are bisyllabic or trisyllabic, with monosyllabic words being avoided (Park, 1995; Myachina, 1981; Maw & Kelly, 1975). Brandon (1975) argues this is because of a rule of penultimate stress that is quite widely adhered to.

- | | | | |
|-----|----|-------------|--------------------|
| (6) | a. | J í ko | fireplace, kitchen |
| | b. | W á tu | people |
| | c. | Ch ú pa | bottle |
| | d. | P í ka | cook (v.) |
| | e. | Kit á nda | bed |
| | f. | Sab ú ni | soap |
| | g. | Ang ú ka | fall (v.) |
| | h. | Ong é a | talk (v.) |
| | i. | Tegem é a | Depend (v.) |
| | j. | Tafadh á li | Please |

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

‘obligation’²⁰), but this is a strong tendency in Swahili, and indeed in most Bantu languages (Kanerva, 1990; Hyman & Katamba, 1990). This stress rule is exemplified below, where primary stress in (7a) is on the verb *pig-*, the penultimate syllable. In (7b), with the addition of the applicative suffix, stress moves rightward onto the applicative suffix, which is now the penultimate syllable. In (8a), stress is on the penultimate syllable of the word *maktába*, a loan word from the Arabic, meaning ‘library’, and with the addition of the locative suffix, stress moves rightward onto the penultimate syllable.

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

²⁰ However, speakers of Nairobi Swahili tend to shift between the standard *lázima* and the more colloquial *lazíma*. Vitale (1985) shows how this shifting between non-penultimate stress and penultimate stress for loan words is common even in standard dialects, suggesting that the shift from non-penultimate stress to penultimate stress is one that takes time. Such words as *lazíma* are new loan words which are in the midst of being incorporated into the phonological system of Swahili, supporting the prominence of the penultimate stress rule in Swahili.

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

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

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

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

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

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

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

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

2.3 Noun Classes

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

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

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

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

- (10) vi – kombe
8–cup
'cups.'

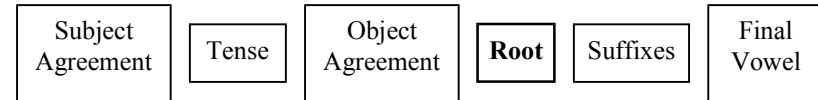
Table 2.1 Meinhoff's Noun Classification System

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

2.4 The Verbal Complex

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

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



Of these elements, subject agreement, tense and the final vowel are the only ones which are obligatorily present with the root in every affirmative Swahili utterance²¹. I will describe the structure, occurrence and function of each of these elements in turn (beginning with Subject agreement and ending with the final vowel).

2.4.1 Subject agreement paradigm

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

(13) exemplify this:

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

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

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

²¹ Although see section 2.4.1 for a description of a non-standard colloquial dropping of SA.

person/number. In example (13), the subject is of noun class #2, and the subject agreement marker changes appropriately²².

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

(14)

	<u>Optional Pronoun</u>	<u>SA</u>	<u>T</u>	<u>V</u>	<u>FV</u>	<u>Gloss</u>
1 st singular	Mimi	ni-	-li-	-anguk-	-a	I fell
2 nd singular	Wewe	u-	-li-	-anguk-	-a	You fell
3 rd singular	Yeye	a-	-li-	-anguk-	-a	He fell
1 st plural	Sisi	tu-	-li-	-anguk-	-a	We fell
2 nd plural	Ninyi	mu-	-li-	-anguk-	-a	You (all) fell
3 rd plural	Hawa ²³	wa-	-li-	-anguk-	-a	They fell

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

2.4.1.1 Nairobi Swahili Vs. Standard Swahili

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

²² In addition to subject verb agreement, these examples show agreement within the subject adjectival phrase. The agreement within the adjectival phrase is agreement between the head noun and the modifying adjective (see Carstens (1991) for a detailed analysis of Swahili DPs and agreement within DPs).

²³ This is specifically Nairobi Swahili. The 3rd person plural pronoun in Standard Swahili is *wao*.

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

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

Table 2.2 Standard Swahili and Nairobi Swahili subject agreement morphology

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

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

2.4.1.2 SA Omission

According to traditional grammars, subject agreement (henceforth SA) is obligatory in all contexts, e.g., Ashton (1947): ‘In Swahili the verb cannot stand alone as in English, but must be prefixed by the Pronominal Concord proper to the noun which forms its subject’ (p.15)²⁴. Similarly,

²⁴ Pronominal Concord = subject agreement

Myachina (1981) claims, ‘The subject markers...are an indispensable component of the verbal complex’ (p.49). Vitale (1981) is just as unequivocal, ‘The important fact is that the subject affix is an obligatory morphological category of the verb’(p.15). Krifka (1995), in his survey of Swahili syntax, is a little more cautious, but nevertheless claims that it is obligatory in all cases except in certain tenses, the habitual *-hu-* and the allomorph of the present tense *-a-* (p.1399). A survey of adult speech in the corpus used in this study reveals that this is not entirely true in this dialect of Swahili. Adult speakers of Nairobi Swahili overwhelmingly use SA in verbal contexts, but occasionally omit SA. This omission does not correspond to the use of habitual *-hu-* or present tense *-a-*, but rather is used in all tense contexts. The discourse context for this omission appears to be in cases when the subject is extremely salient, and when the topic and the subject are co-referent. Of all verbal utterances in the adults’ speech, 5% of their utterances were missing SA. While this is a small proportion of omitted SA markers, as I will discuss later in the syntax section of this chapter, omission of SA is considerably more frequent than omission of any other obligatory affix, for example, tense or mood, omission of which is nonexistent. So the omission of SA constitutes a genuine grammatical possibility for adults²⁵. An analysis of this sentence type will be provided later in this chapter.

²⁵ In fact Ashton (1947) notes that SA may be omitted in certain clauses that take the *ka* tense marker. She claims that the omission of SA results in an emphatic or surprised connotation. Scotton (1969) also describes cases in which up-country Swahili speakers and Bagandan Swahili speakers frequently omit SA. See below for details.

2.4.2 Tense/Aspect

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

Table 2.3 Tense/aspect morphemes in Swahili

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

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

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

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

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

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

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

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

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

(21) Ku - imb-a ku - me - kwish - a → Gerundive Infinitive
 INF -sing -IND SA_{1s}-pr.prf.-end - IND
 'The singing has ended.'

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

2.4.3 Negation

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

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

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

In this case, the only overt representation of negation is the negative marker at the beginning of the verbal complex. However, in other tenses the two other negative positions are activated. Table 2.4 below presents the negation paradigm²⁶:

Table 2.4 Swahili Negation Paradigm

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

In the case of the past *li*, the negative form is a negative correlative, i.e., *hu*, appears before the SA marker and the suppletive form *ku* of the tense marker *li* occurs in tense position.²⁷ For the present tense, when *ha* appears before SA, we find the verb marked with a negative final vowel. As

²⁶ The tenses which are not shown in this table (narrative, hypothetical, habitual/conditional, and infinitive) form their negation through a complex negation. A cleft-construction is used which clefts an auxiliary *be* verb plus negation, and this forms the only mechanism for negation in these circumstances. An appropriate translation of such a construction would be "It was not that..."

²⁷ Note that this negation particle is homophonous with both the 2nd person singular Object Agreement Marker as well as the Infinitive marker.

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

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

2.4.4 Object Agreement Marking

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

The Object agreement marker is obligatory in simple transitive sentences when the object is specific and/or animate²⁸. When the object is non-specific and inanimate, object agreement is obligatorily absent. We will return to this point in section 2.7.

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

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

Table 2.6 Personal pronoun Object Agreement

Person	Object Agreement
1 st singular	ni
2 nd singular	ku
3 rd singular	m
1 st plural	tu
2 nd plural	mu ²⁹
3 rd plural	wa

²⁸ See Bresnan & Mchombo (1987) for a discussion of the object agreement marker in Bantu, and Moshi (1985) and Ngonyani (1996) for a specific discussion of object agreement in Kiswahili.

²⁹ In Standard Swahili, the OA for 2nd plural is *wa*, not *mu*.

- (23) ni – ta – **zi** – nunu– a nguo
SA_{1s}–fut–OA₁₀–buy–IND clothes
'I will buy the clothes.'
- (24) A – li – **ni** – pig - a
SA_{3s}–past–OA_{1s}–hit–IND
'He hit me.'
- (25) A – li – **ku** – pig - a
SA_{3s}–past–OA_{2s}–hit–IND
'He hit you.'
- (26) A – li – **m** – pig - a
SA_{3s}–past–OA_{3s}–hit–IND
'He hit him'³⁰

2.4.5 Roots

The verbal root in Swahili, and indeed in most Bantu languages, is generally monosyllabic. Not all verb roots are monosyllabic however, and the majority of the multi-syllabic verb roots are verbs which were originally borrowed from other languages (most commonly from Arabic). The verb root is generally CVC, although other structures are not uncommon³¹:

³⁰ The reference of the object him cannot be the subject him, i.e., this cannot take an anaphoric interpretation. For an anaphoric interpretation the anaphoric marker *-ji-* must be used in preverbal position instead of the 3rd person singular marker *-m-*.

³¹ In this table, the roots are given along with the final vowel, which in Swahili is obligatory.

Table 2.7 Some root structures for Swahili verb roots

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

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

Table 2.8 monosyllabic verb roots and their citation forms

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

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

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

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

However, these monosyllabic verbs have been reanalyzed in Nairobi

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

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

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

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

2.4.6 Suffixes

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

Table 2.9 Common Verbal Suffixes in Swahili

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

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

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

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

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

Table 2.10 Types of Applicative sentences in Swahili

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

The example (30) shows a non-applicative sentence, and (31) shows its applicative counterpart. (31) is a benefactive applicative, while (32) is an instrumental applicative and (33) is a locative applicative³².

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

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

³² The locative example is Ngonyani's example, and so the verb root is given as [l]. In Nairobi Swahili the verb has incorporated the dummy infinitive into the verb root, resulting in [kul].

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

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

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

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

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

The final suffix to be exemplified is the reciprocal suffix, as in (39).

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

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

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

2.4.7 Final Vowel

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

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

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

Table 2.11 Contexts for use of final vowel

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

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

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

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

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

2.4.8 Imperatives

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

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

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

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

In example (44) there is an object agreement marker, which in Nairobi Swahili is required by the specific applicative object (see section 2.4.6 on the applicative)³³. In cases where there is no applicative, no object agreement is required, and so the imperative form surfaces as a bare stem.

³³ OA is usually absent in imperatives in Standard Swahili.

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

- **Syntax of Swahili**

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

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

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

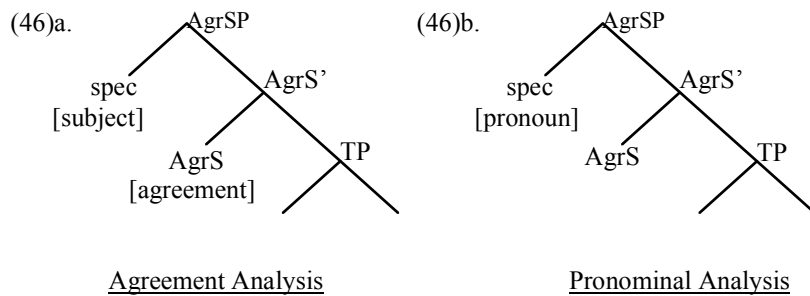
2.5 Subject Agreement versus Subject Pronoun

A current debate in Bantu linguistics centers on whether SA is an agreement marker or a pronominal clitic (e.g., Bresnan & Mchombo, 1987; Keach, 1995; Zwart, 1997)³⁴. Traditionally, it has been described as agreement (Ashton, 1947; Myachina, 1981), although some of its functions are pronominal in nature. I will discuss the two sides of this debate and determine the nature of SA in Nairobi Swahili. While the importance of this will become clear once we begin discussing the acquisition data, a priori we have reason to pay attention to the nature of SA. We saw in

³⁴ Similar debates are currently going on for other languages that have rich agreement morphology, e.g., Spanish (Ordoñez, 1997; Goodall, 2002; Grinstead, 1999), Tongan (Otsuka, 2001), etc. Furthermore, there are dialects of English in which preverbal pronouns are analyzed as agreement markers, e.g., Bošković & Chapman (1998) show that in certain non-standard dialects in English, the preverbal pronoun acts more like a subject agreement marker than a pronominal subject, e.g., the ‘pronoun’ is bound to the verb stem, conjunction is impossible, null subjects are licensed, etc. They argue that English pronouns are in the midst of a move from pronouns to agreement clitics (as suggested by Givón, 1975). We will see that Swahili may be in the same process.

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

Let me begin by clarifying the difference between an ‘agreement analysis’ and a ‘pronominal analysis’. Agreement is a process in which two elements that are in a local configuration (46a)³⁵ share morphological features through a process of feature matching. Although not crucial to my analysis, I assume that the subject raises from a lower VP position to [spec, AgrSP] (Koopman & Sportiche, 1991) and triggers agreement with the head AgrS. This subject can be optionally null, in which case it is licensed through rich agreement (Rizzi, 1982; Taraldsen, 1978; see section 2.5.1).



³⁵ Recent work in Minimalism (Chomsky 1998, 2001) has raised the possibility that agreement relations need not be strictly local, as in 38a. I will not consider the possibility of a slightly ‘looser’ agreement configuration because all agreement relations are strictly local in Swahili. In other languages in which non-local relationships occur, we might need to reconsider our understanding of agreement.

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

Bresnan & Mchombo (1987) argue that in Chiche_a, an East Central African Bantu language, SA is ambiguous between agreement and a pronoun. They use several functional diagnostics, including the interplay of word order with agreement morphology and tone with phrase structure. They show that overt subject NPs differ from overt object NPs in their distribution in relative clauses and interrogatives, these being contexts which allow and disallow topics respectively (see the next section for more detailed reasons for why this is so). Chiche_a differs from Swahili in some significant ways,³⁶ and so it does not benefit us to discuss the details of the

³⁶ For example, Chiche_a has tone, Swahili does not. Chiche_a does not exhibit the ‘Definiteness Effect’ in Object Agreement (Bresnan & Mchombo, 1987, p.761), while Swahili does. Chiche_a allows postverbal subjects to be questioned, a fact that Bresnan & Mchombo take to mean that ‘the subject and topic NPs appear at the same level of structure in the S, with exactly the same ordering possibilities’ (p.775). Swahili, on the other hand, disallows such questions:

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

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

2.5.1 Keach (1995)

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

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

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

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

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

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

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

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

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

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

can receive its theta role directly (example 40a)³⁷. For Keach, a DP in topic position is sufficiently local for theta assignment to occur, but once that DP is moved (either postposed or raised to a higher clause), that local relationship no longer exists³⁸. According to Keach, it is unclear how to

³⁷ Keach provides examples which show that subjects in *Hu*-clauses must be overt:

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

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

³⁸ As discussed earlier, according to Keach, in the case of non-habitual clauses the theta role can be transmitted through a chain from the subject pronoun to the topic DP. However, it is not clear why in *hu*-habitual clauses, a theta role cannot be assigned to the trace of the topic and transmitted by the chain that connects it to the moved topic. This cannot be due to the different A/A' status of these chains, as both are A'-chains. Thus I argue the ungrammaticality of postposed topics in *hu* habitual clauses is not due to the status of SA as a pronoun.

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

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

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

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

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

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

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

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

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

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

2.5.2 Additional arguments for agreement

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

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

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

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

Thus the restriction on quantified topics holds in Swahili³⁹. Under a pronominal analysis of SA, the preverbal DP is in topic position, and so a quantifier should be ungrammatical. However, as (56) shows, quantifiers are possible in preverbal position, suggesting that the preverbal DP is in subject position.

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

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

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

In Swahili, the preverbal DP can be the answer to a question⁴⁰:

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

⁴⁰ Thanks to Stan Dubinsky and Ivano Caponigro for discussions on this point.

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

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

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

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

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

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

taken to be prefixes rather than (pronouns)” (Steele, 1995, p.1227)⁴¹. In making this distinction, Steele (along with Jacobs, 1975; Steele, 1977; Langacker, 1977) uses word order as a diagnostic for whether a subject marker is an agreement affix or a pronominal clitic, with the former fixed in position with respect to the verb, and the latter being somewhat freer.

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

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

⁴¹ The only examples she gives are to illustrate that these prefixes combine subject and object marking, and not to illustrate the unacceptability of free word order. One example is her example (7a):

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

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

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

2.6 Tense versus Auxiliary Verb

For the same reasons we investigated SA in adult Swahili, we must investigate tense. Children acquiring Bare verb languages or RI languages tend to either omit inflectional affixes or use infinitival morphology, suggesting that tense is somehow a late acquisition. It is important for us to understand the properties of tense in adult Swahili in order to better understand the acquisition data. Traditionally, the tense marker has been described as an inflectional prefix. Recently, this view has been challenged, and the tense marker has been analyzed as an auxiliary verb. Under this view, the verbal complex is composed of the auxiliary (tense) verb with a SA prefix, and the main verb with OA prefix. There are several reasons why tense may be viewed as an auxiliary. I will present the evidence for this position, along with criticisms. First, many tense markers are transparently related to lexical items that are clearly verbal (Zwart, 1997). Examples are shown in (60).⁴² The argument is that these elements were originally verbs that have developed into auxiliary verbs.

(60)

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

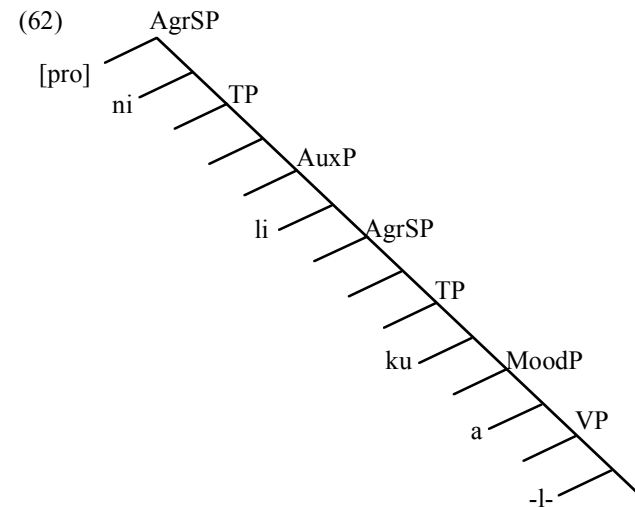
⁴² The transparency is obvious for *-na-* and *-ta-* but not so for *-me-*.

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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

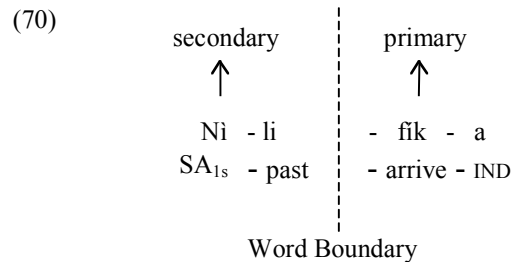
Such constructions occur only in Standard Swahili, not Nairobi Swahili and so do not bear on this study. In Nairobi Swahili coordination occurs either with the verb stem, as in (68), or with the entire verbal complex, as in (69).

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

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

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

The final argument for tense being analyzed as an auxiliary is the stress pattern exhibited in the verbal complex. As noted in section 2.2, Swahili has a rule of penultimate stress. In a simple verbal complex, this occurs on the main verb. However, secondary stress is also present, usually on the SA marker. Barrett-Keach (1986) interprets the secondary stress as evidence that there is a word boundary between tense and the rest of the verbal complex⁴³. This word boundary may suggest that tense is a separate lexical item, with SA being a prefix to this verb (Zwart, 1997).



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

⁴³ She doesn’t actually argue for an auxiliary analysis of tense, but is more concerned with showing that there is a syntactic constituent Aux, which is manifested in Swahili as SA+T, and is a distinct word.

- (71) a. Ni – me - fík – a
 SA_{1s} – perf. – arrive – IND
 ‘I have arrived.’
- b. Ni – mēsha – fík – a
 SA_{1s}–perf.comp.–arrive–IND
 ‘I have already arrived.’
- c. Ni – mekwisha – fík – a
 SA_{1s}–perf.finish–arrive–IND
 ‘I have finished arriving.’

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

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

and the verb in Swahili. Thus the word boundary that exists between tense and the verb is clearly not loose enough to allow intervening adverbs.⁴⁴

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

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

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

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

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

⁴⁴ Furthermore, Zwart (1997) shows that in some relative constructions in which subject inversion is possible, you would expect the subject to intervene between the ‘auxiliary’ and the main verb. However this is impossible, with the only grammatical inverted structure being the post-main-verb construction. Zwart concludes that ‘the auxiliary and the main verb do form a unit of some kind’.

2.7 Object Agreement versus Incorporated Pronoun

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

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

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

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

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

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

(Keach, 1995, examples 12a-d)

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

There are two problems with this argument. First, according to Keach, in every other context of agreement in Swahili, the agreement marker is obligatory⁴⁵: SA is obligatory with both animate as well as inanimate objects, and OA is obligatory with animate objects. It is unclear why OA is ‘optional’ with inanimate objects. Secondly, the claim that OA is obligatory with animate objects is not correct. Keach fails to mention the

⁴⁵ Although see section 2.4.1 for exceptions that Keach does not mention, namely the [-SA] clauses.

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

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

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

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

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

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

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

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

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

The second objection arises from the data itself. The examples in (78)-(81) all make use of a past tense marker, indicating that the event that is being denoted is complete and known to the speaker (in its entirety). Therefore, this pragmatically forces a specific reading of *watoto* ‘children’⁴⁶, hence requiring OA. If we change the tense marker to *na* present tense, which allows for a habitual/generic reading, we can eliminate this bias.

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

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

⁴⁶ A non-specific, past complete reading would require a very complex context: He used to like children, but then something happened and he doesn’t like children anymore. While this is possible, without this context being explicitly constructed, such an interpretation is extremely unlikely.

animate objects obligatorily require OA, and so they will reject the sentence on that basis⁴⁷. Speakers who reject the non-inverted version of the sentence (mwaliimu anapenda watoto) because of the lack of OA will also reject the sentence in (82). The examples that Keach provides show that her consultants were such speakers, and so it is not surprising that Keach was led to the conclusions she reached. Second, a topicalized object tends to be previously-mentioned information, and so is usually specific. A non-specific reading such as that in (82) is possible, but requires a very carefully created context.

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

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

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

⁴⁷ I am basing this on my own experiences, as well as reports from all my native speaker consultants.

agreement, not a pronoun.⁴⁸

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

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

⁴⁸ In addition to the idiomatic objects argument, she uses interrogatives to show that inanimate OA is pronominal and animate OA is optionally agreement or pronominal. Her examples are:

- | | | |
|---|--|------------------|
| a. a – li – on – a nini? | b. *a – li – ki – on – a nini? | Inanimate |
| SA _{3s} -past-see-IND what | SA _{3s} -past-OA ₇ -see-IND what | |
| ‘What did he see?’ | | |
| c. a – li – mw – on – a nani? | d. *a – li – on – a nani? | Animate |
| SA _{3s} -past-OA _{3s} -see-IND what | SA _{3s} -past-see-IND what | |
| ‘Who did he see?’ | | |

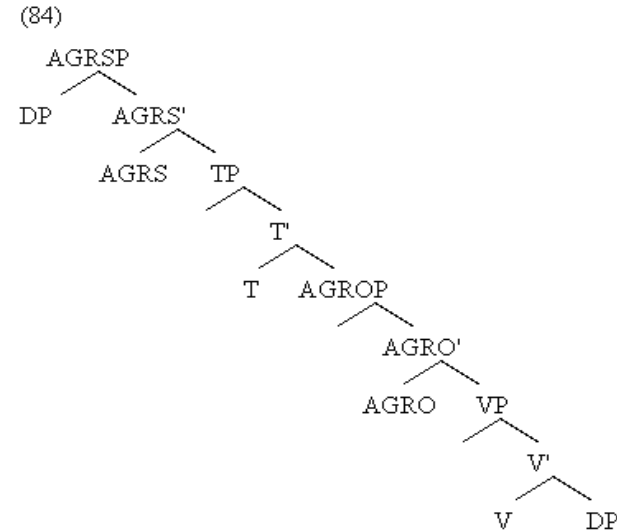
These examples show that OA is possible in interrogatives only when the direct object is animate (ex.c-d). This does support Keach’s claim that OA is pronominal with inanimate objects but can be agreement with animate objects. However, this data is incompatible with the data presented in the text. Reconciling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

applicative, in Swahili and Ndendeule, see Ngonyani (1996). These remain important and interesting avenues for further research.

2.8 The Tensed Clause

Traditionally, the Bantu verbal complex has been analyzed as one large macrostem with smaller morphemes attached onto it. Bantu languages are generally seen to be quite homogeneous with respect to the syntactic phenomena they exhibit. This is especially true of the Eastern Bantu languages (Wald, 1990). Given this, I will use evidence from various Bantu languages, in order to formulate a syntactic analysis of the Swahili functional structure.

I will describe the basic syntax that I assume for Swahili, although little hinges directly on this particular analysis. Most of what follows in subsequent chapters is compatible with alternative analyses and frameworks. Following Pollock (1989) and Chomsky (1989), Demuth & Gruber (1995) suggest a structure for Sesotho (a South Eastern Bantu language spoken in Lesotho and South Africa) with a split INFL, one which includes AGRS, T and AGRO.



Under this analysis, the subject originates within the VP (Koopman & Sportiche, 1991), and raises to Spec-AGRSP. The verb raises through AGRO, T and into AGRS. It is also proposed that *pro* is licensed in a A-position [Spec, IP], and as in Italian, *pro* is identified by rich agreement. Therefore, Sesotho is identified as a typical *pro*-drop language. Swahili is very similar to Sesotho in this respect, and as the preverbal structure of Swahili is virtually identical to Sesotho, let us adopt this structure as a first approximation.

In fact, Ngonyani (1996), using evidence from various sentence types (including applicatives, negative sentences, adverbials, etc), adopts a similar structure for Swahili, as well as Ndendeule (a closely related Bantu language spoken in southern Tanzania). He proposes a structure similar to that proposed by Demuth & Gruber, but he adds to it a clitic phrase between

TP and AGROP (see below for discussion). Furthermore, he provides extensive evidence for verb raising. He shows that when VP ellipsis occurs, the verb itself is not elided, since it has raised out of the VP.

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

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

object within the VP to be elided⁴⁹.

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

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

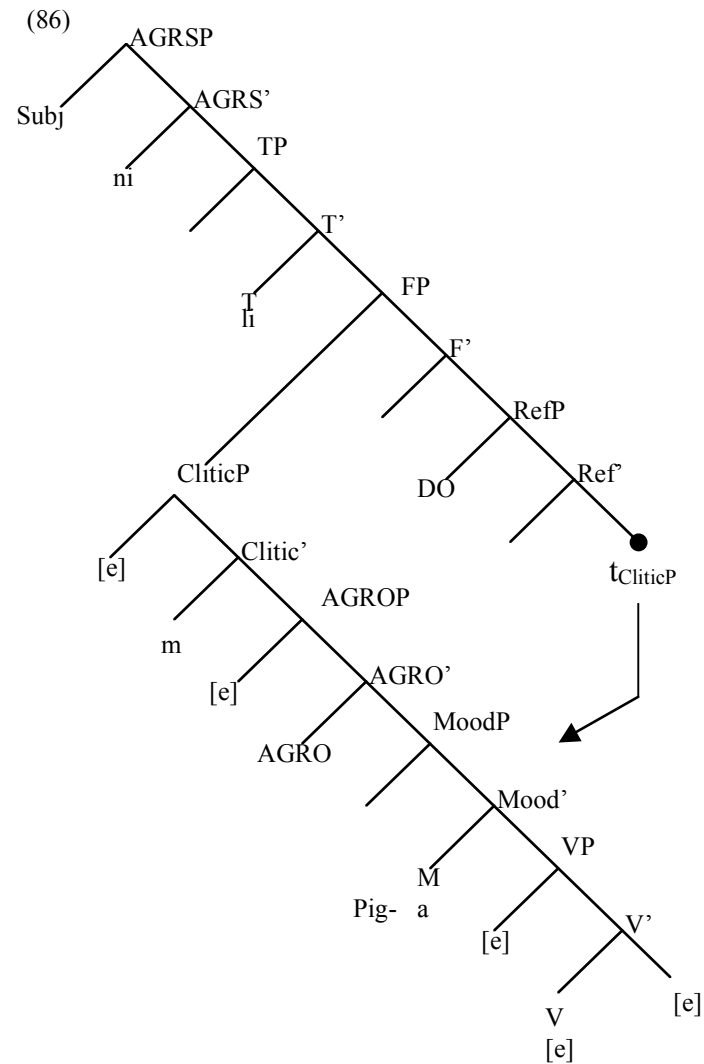
⁴⁹ As an aside, Ngonyani also shows quite convincingly that the applicative structure contains a Larsonian VP shell, with the higher VP headed by the applicative morpheme, and the lower VP headed by the verbal root. Each VP has a separate AGROP. He finds that when VP ellipsis occurs, the verb remains intact (hence, he concludes that V raising has occurred prior to ellipsis). Furthermore, when ellipsis occurs in double object constructions such as the applicative, either both objects are elided or only the lowest object is elided. It is never the case that the left most object is elided with the right most object remaining intact. This shows evidence of a hierarchical structure within the VP.

are checked, and OA is licensed in the syntax⁵⁰. The lexical subject, which originates in [spec, VP] raises to [spec, TP] to check case features, and then to [spec, AGRSP] to license the SA. I further assume that the verb also raises and left adjoins to the mood marker which is in the head of MoodP (immediately dominating VP). This gives the correct order of V-mood and is consistent with Ngonyani's ellipsis facts.

Given these movements, the morpheme order in the derivation still does not match the observed morpheme order in adult Swahili. Specifically, the morpheme order at this point is the following:

Subject SA T DO OA V Mood

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



There are several points that are worth highlighting about this analysis. First, the analysis involves both phrasal movement as well as head movement. Second, the final phrasal movement is an instance of remnant

⁵⁰ The DO actually raises further, to a projection which dominates CliticP. This move has two reasons: first, without such a move, we would have a violation of a principle of syntax which derives from the doubly-filled COMP filter. Namely, the configuration of having an overt head and an overt spec is ungrammatical. The second reason is that the DO is a referential element, and so it moves to the spec of RefP (cf. Beghelli & Stowell, 1992). For details, see Deen (1999).

movement (Webelhuth, 1992; Koopman & Szabolci, 1998). Remnant movement occurs when a phrase which contains the trace of an already moved element moves to a position in which it c-commands the previously moved element. For justification of this type of movement see Koopman & Szabolci (2000) and references therein. Third, there is a branching point after the Tense. This may be related to the stress facts we saw earlier in which the right edge of T acts as a phonological word-boundary. Some theories of the syntax-phonology interface posit that crucial branching points (or XP boundaries) in the syntactic structure are mapped onto phonological breaks in the phrase (e.g., Selkirk, 1986). This is not incontrovertible evidence for such a position, but is simply suggestive. And finally, the verbal complex does not constitute a complex head in the syntax. Rather, each morpheme occurs distinctly, with the full morphological verbal complex being created at SPELLOUT (see Julien, 2000 and references therein).

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

In this chapter so far we have seen that SA is best analyzed as agreement between the subject and the verb, as is OA. I argued that while the evidence for tense being an inflectional clitic rather than an auxiliary verb is less convincing it is nevertheless the best analysis, given the available evidence. We will return to these issues in chapter 4 when we look at child Swahili. In the next section I will discuss null subjects in adult

Swahili, and show that the null subject has the characteristics of *pro*, and resembles Italian null subjects in important ways. I will conclude that it is in fact *pro*. In the following section, 2.10, I will introduce a class of clauses that all allow SA omission.

2.9 Null Subjects – null *pro*

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

- | | | |
|------|--|---------|
| (87) | *He said that [e] is eating | English |
| (88) | Ha detto che [e] mangia
have-3s said that eat-3s
'He said that (he) is eating' | Italian |

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

The fact that null subjects occur in finite clauses in Italian means that the null element is not PRO (as PRO only occurs in non-finite contexts). Furthermore, the null subject in the example above has a definite specific reference, as opposed to an antecedent-controlled reference (as with PRO) or an arbitrary reference (as with PRO_{arb}). Null subjects are thus more akin

to overt pronominals. Chomsky (1982) concludes that the null element in subject position in a clause such as (88) is the null counterpart to regular pronouns, and is called *pro*. The distribution of overt pronouns and *pro*, however, are different. For example, in Italian *pro* cannot occur as the object of a preposition, but overt pronouns can:

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

(90) Ho parlato con lui Italian
 have-1s spoken with him

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

(91) a. ma [pro] w_-xwar-a Pashto
 I PRF-eat-OA_{fem-sg}
 'I ate (it-fem)'

b. * z_ [pro] xwr - _m Pashto
 I eat - SA_{1st-masc}
 'I eat (it)'

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

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

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

(93) a. *[e] will zu Hause bleiben German
 want at home to-stay
 '(I) want to stay home'

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

These facts have generally been interpreted as pointing to the existence of two distinct conditions on null subjects: a licensing condition, and an identification condition (Rizzi, 1986). The licensing condition applies to all null pronouns, while the identification requirement only applies to referential/argumental null pronouns. The licensing requirement states that a *pro* must be licensed by its governing head. In Minimalist terms this can be interpreted as *pro* having Case features that must be checked. As for identification, in order for a noun to be referential, it must be specified for person / number features. Therefore, the identification requirement states that a referential pronoun must get Φ -features through co-indexation with a case-governing head. In Minimalist terms, the pronoun must have its Φ -features checked by an appropriate head. Presumably there is a relation between rich agreement and the existence of Φ -features on that head in order to allow identification, though the exact specification of "rich

agreement” remains elusive. Therefore in Italian *pro* is identified because the language has rich agreement, while in English this is not the case.

(94) *pro* Parl-o Italiano Italian
 └───┬───┘
 identification

(95) **pro* speak English English
 └───┬───┘
 no identification possible

How does this solve the problem raised by German? Rizzi claims that German satisfies the licensing requirement, but not the identification requirement. Thus, German licenses non-referential null pronouns, but because it does not satisfy the identification requirement, null referential pronouns are disallowed.⁵¹

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

⁵¹ Additionally, there are languages such as Chinese that allow subject and object omission without any agreement whatsoever. This is a problem for Rizzi’s proposal. Huang (1984) proposes that these are variables bound by a null topic operator. We shall return to this point in a later section when examining Swahili [-SA] clauses.

from PRO. For example, both Swahili null subjects and Italian *pro* alternate with overt DPs:

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

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

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

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

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

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

2.10 Subject Agreement Omission

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

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

(99) SA – T – V - IND

However, to my knowledge there have been no corpus-based studies that examine the question of whether SA may ever be omitted.⁵² In the dialect of Swahili spoken by the parents of the children in this study, SA omission is noticeably frequent (see below for more details). Furthermore, my native consultant and I agree that SA omission is grammatical, but only under certain conditions. In what follows, I will describe some of these conditions.

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

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

⁵² Carol Meyers-Scotton (p.c.) informs me that SA omission is fairly frequent in the spoken forms of many dialects of Swahili. In fact, she documents such a phenomenon (Scotton, 1969) in the dialects of Baganda and Baluhya speakers in the 1960s. She finds that they frequently omit SA prefixes and rarely use incorrect SA (p. 106). She gives examples such as the following, in which the first line is the dialect, and the second line indicates the Standard Swahili equivalent:

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

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

2.10.1 Imperatives

(100a) below is an imperative, in which SA is obligatorily absent

(note the ungrammaticality of 100b).

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

Cross-linguistically, imperatives are often (although not necessarily) unmarked for subject agreement (see Koopman, 1997). This appears to be a quite general property of imperatives, an explanation of which goes beyond the scope of this dissertation (see Potsdam, 1995; Platzack & Rosengren, 1997; Zanuttini, 1997; Koopman, 1997 for details). I will discuss the other three types of clauses: infinitives, habituals and continuative clauses, and later extend the analysis of continuative clauses to a general phenomenon of SA omission. I will then describe three significant differences between habituals and continuative clauses. I will use these differences to argue that the syntactic structures associated with these two clause types are significantly different.

2.10.2 Infinitives

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

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

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

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

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

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

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

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

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

2.10.3 Habituals

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

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

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

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

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

2.10.4 Continuative Clauses

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

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

- b. * a – li – ka – kimbi – a
 SA_{3s}–past–cont–run–IND
- c. * a – ka – li – kimbi – a
 SA_{3s}–cont–past–run–IND

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

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

2.10.5 Differences between Habituals and Continuatives

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

2.10.5.1 Optionality of SA

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

2.10.5.2 Optionality of subjects

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

- (111) a. ulevi hu – ondo – a akili
 drunkenness HAB–remove–IND sense
 ‘drunkenness removes common sense’
- b. * hu – ondo – a akili (Examples cited in Keach, 1995)
 HAB–remove–IND sense

However, the subject of a continuative may be either null or overt. In the unmarked case (when the subject of the continuative clause is the same as the subject of the previous discourse), the subject is null. However, the subject may be overt when there is a change in subject or a clarification required. For example, in (112a), the subject of the second (continuative) clause is the same as the subject of the main clause. Similarly, in (112b) the subject of the second clause must be the same as the subject of the first clause if the subject is null. When the subject of the second clause is not identical to the subject of the first clause, as in (112c), then an overt subject is required⁵³.

⁵³ While our eventual goal is to understand SA omission, I am exemplifying here the fact that subjects may be overt or null in continuative clauses in general. This same fact is true in continuative clauses that are missing SA.

(112)

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

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

2.10.5.3 Embedding

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

- (113) a – li – ni – ambi – a [kwamba wa – tu wa Kenya
 SA_{3s}–past–OA_{1s}–tell–IND that 2-person of Kenya

 hu – wa – pend – a wa – toto]
 HAB–OA₂–like–IND 2-child

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

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

- b. ?? a – li – ni – ambi–a [kwamba Ø ka – kimbi – a]⁵⁴
 SA_{3s}–past–OA_{1s}–tell–IND that Ø cont – run – IND
 ‘He told me that (he) then ran off’

The differences that we have seen so far are summarized in table 2.12:

Table 2.12 Summary of characteristics of habitual and continuative clauses

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

Recall from the discussion in 2.9 that null subjects must be identified (Rizzi, 1982; Jaeggli & Safir, 1989). Identification can occur through several means, such as control (in the case of PRO), the presence of a c-commanding antecedent (in the case of a trace), or through rich agreement features on a licensing head. This rich agreement provides an identifier for null *pro*. We saw earlier in this chapter that the null subject of a Swahili tensed clause is *pro*. However, in the case of habitual and [-SA] continuative clauses, this rich agreement is missing. Therefore the question arises as to what the status of the null subject is in such clauses. Specifically, given that the null subject occurs in the absence of SA, how is the identification requirement satisfied?

We saw in the examples in (111) and the summary in table 2.12 that habituals simply do not allow null subjects. Therefore, the answer to the question for habituals is clear: because rich agreement is absent, null subjects are blocked. This is consistent with our theory of identification of

⁵⁴ My consultant considers this sentence ungrammatical. My judgment is somewhat less clear, but certainly degraded.

null *pro*⁵⁵. I therefore assume that subjects in habitual constructions are structurally in subject position and must be overt because of the lack of an identifier. However, continuative clauses allow the omission of SA, and in those same clauses a null subject is possible. This is not expected under our theory of identification. Additionally, the fact that [-SA] continuative clauses are not possible in embedded contexts is surprising (cf. examples 114), as embedded *pro* clauses are possible in Italian, as well as in Swahili full clauses:

- (115) *pro* so che cosa *pro* hai detto Italian
 know-1sts what thing have-2nds said
 ‘(I) know what (you) said’

- (116) *pro* ni – na – ju – a *pro* u–li–sem–a nini Swahili
 SA_{1s}–pres–know–IND SA_{2s}–past–say–IND what Full Clause
 ‘(I) know what (you) said.’

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

⁵⁵ This is an alternative view of the facts presented by Keach (1995) in section 2.5.1. Recall that habituals do not allow postverbal subjects. Keach argues this is because SA is a pronoun, and in the absence of this pronoun or a local (preverbal) subject, $_$ -role assignment is impossible in habitual clauses. This may well be true, but another view is that the presence of rich agreement allows a free word order. In Italian (a language with rich agreement), postverbal subjects are allowed. However, in English (a language with no rich agreement), postverbal subjects are not allowed. Therefore agreement seems to be important in allowing postverbal subjects, not necessarily the existence of a clitic pronominal subject, as Keach argues.

of empty categories occur in the input language. Thus we will investigate the omission of SA and the properties of the null subject in such clauses.

Unfortunately the descriptive evidence available in the literature as to when SA may be omitted is very limited. In order to gain a better empirical understanding of SA omission, I investigated the use of SA by the adults in the Swahili corpus. The first thing I looked at were habitual clauses: not a single utterance containing the *hu-* prefix occurred in the entire corpus. Secondly, there were also no cases of continuative *ka*. This is most likely due to the context of the recordings. The continuative *ka* is used to tell stories, and the purpose of these recordings was to elicit stories (or any speech) from the children. Thus, continuative *ka* never occurred in the recordings. However, I discovered that SA was omitted in a significant proportion of adult speech in non-continuative contexts. In the next section, I will discuss the contexts of these [-SA] clauses in Nairobi Swahili. I will then provide an analysis of these clauses which postulates a null constant (Rizzi, 1992; 1997) as the null element in subject position. Later in chapter 5 I will extend this analysis to the speech of children, showing that this analysis sheds light on other underspecified forms in child speech.

2.11 [-SA] Clauses

I conducted a CLAN analysis on the Swahili corpus targeting the adult utterances in 16 files sampled from all four children. I investigated the omission of SA, the expression of tense in these clauses, the identity of the missing referent, and the occurrence of overt subjects in these clauses. Most of the examples that I will provide come from the spontaneous speech of the adult speakers in the Swahili corpus. However, all examples have

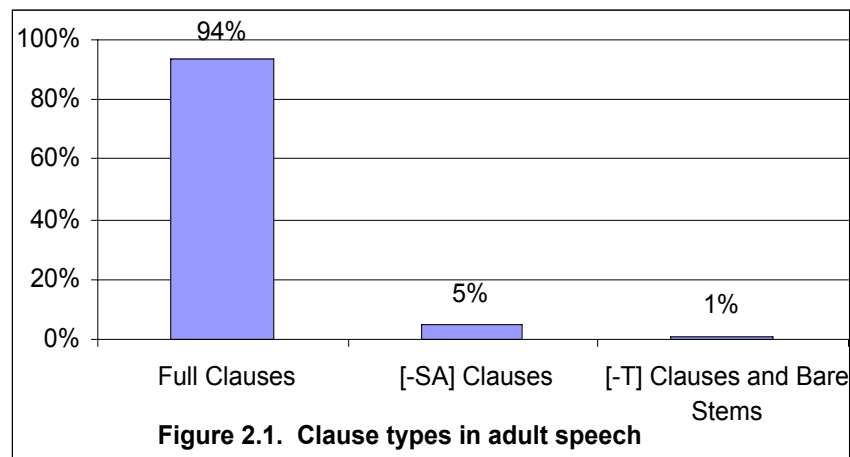
been verified with my native consultant (as well as my own judgments), and differences in judgments are noted.

2.11.1 Frequency of [-SA] Clauses

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

Table 2.13 Proportions of different clause types in adult Swahili.

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



2.11.2 Tense in [-SA] Clauses

[-SA] clauses occur with a variety of tense markers⁵⁶:

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

2.11.3 Implicit Reference of [-SA] Clauses

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

- | | | |
|-------|---|--|
| (120) | Ø ta – ku – chapa – a
fut-OA _{2s} – slap-IND
'(I) will slap you' | 1 st singular
(Sam, MUS10) |
| (121) | Ø na – ju – a ku-wach- a kelele ?
pres-know-IND inf-leave-IND noise
'(Do you) know how to stop making noise?' | 2 nd singular
(Ala, MUS09) |
| (120) | n-dege Ø na – ruk – a
9-bird pres-fly up-IND
'The bird is flying up' | 3 rd singular
(Ali, FAU07) |

⁵⁶ None of the adults used the past tense marker in [-SA] clauses in this corpus. However, my consultant considers the past tense in a [-SA] clause grammatical, and in child speech the past tense marker is used on several occasions.

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

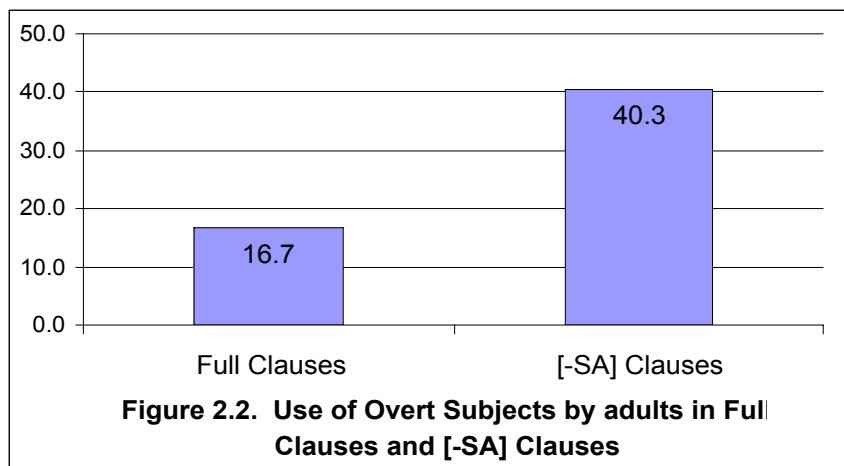
2.11.4 Overt Subjects in [-SA] Clauses

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

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

- | | | |
|-------|--|--|
| (123) | a. wewe Ø ta-kul – a ch-akula?
You fut-eat-IND 7-food
'Will you eat food?' | Overt Subject
(Ala, MUS08, line 230) |
| | b. ndege Ø na – ruk – a
bird pres-climb-IND
'The bird is climbing.' | Overt Subject
(Ala, MUS12, line 2372) |
| (124) | a. ndio, Ø ta – i – beb – a
yes fut-OA-carry-IND
'Yes, (I) will carry it.' | Null Subject
(Ali, FAU01, line 178) |

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



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

2.12 Null Subjects in [-SA] Clauses

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

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

These clauses occur relatively infrequently when compared to full clauses, but when they do occur, they occur primarily with null subjects. We conclude that the null subject is not *pro*, but some other null element

that receives identification through some means other than agreement. This null element is syntactically active, as seen in the following [-SA] examples. In (126), the null subject is the antecedent to the reflexive prefix *-ji-*. In (126), the null subject is the controller for the embedded PRO⁵⁷.

- (126) Ø na – ji – on – a
 pres – refl. – see – IND
 ‘(I) see myself.’
- (127) Ø_i na – ju – a PRO_i ku – onge – a?
 pres – know – IND inf – speak – IND
 ‘Do (you) know how to speak?’

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

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

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

We saw earlier that PRO occurs in non-finite clauses in Swahili. However, we can rule PRO out from [-SA] clauses for three reasons. First, PRO occurs prototypically in embedded clauses, and as we saw in (125c), [-SA]

⁵⁷ The reflexive example is a constructed example, and the PRO example is an actual utterance from the Swahili corpus (MUS09, line 131).

clauses do not occur as embedded clauses. Second, PRO occurs in tenseless clauses, while [-SA] clauses always occur with Tense (cf. Examples 117-

119). Third, PRO does not usually alternate with overt DPs:

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

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

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

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

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

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

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

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

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

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

[-SA] clauses quantified antecedents are either ungrammatical or marginal at best⁵⁸:

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

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

Summarizing, we have found that the null element in subject position of a [-SA] clause cannot be *pro* (no identifier), it cannot be PRO (doesn’t occur in embedded clauses), it cannot be an NP trace (NP traces do not alternate with overt DPs), and it cannot be a wh- trace (it cannot be bound by a quantified antecedent).

2.13 Rizzi’s Null Constant

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

⁵⁸ There is variation in judgments on this point, as my consultant disallows all quantificational antecedents to [-SA] clauses, but I find wh- antecedents marginal and other quantifiers ungrammatical. An additional test for a variable is whether it is sensitive to weak crossover effects (Lasnik & Stowell, 1991). Swahili does not have wh- movement, and so this is difficult to test.

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

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

(135) $[_{TopP} OP [_{IP} nc [_{VP} \dots]]]$

|

|

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

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

(136) a. *(Ich) habe es gestern gekauft* German
 '(I) have it yesterday bought'

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

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

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

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

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

However, Rizzi notes a fact first pointed out by Cardinaletti (1991), that in colloquial German there is an asymmetry between subject drop and object drop. Cardinaletti claims that 'subject drop can involve pronouns of any specification, while object drop is restricted to 3rd person'. She claims that 3rd person specification is a property inherent to operators, and concludes that subjects should not include an operator. Rizzi therefore limits the above structure (135) to object omission, and proposes that the structure for German subject omission is as follows:

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

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

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

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

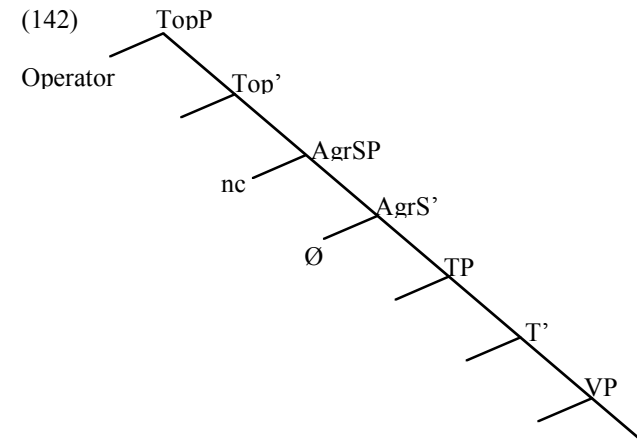
The structure in (139) violates the ECP as stated above, but Rizzi proposes an addition to the ECP:

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

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

2.14 Null Constants in Swahili

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



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

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

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

2.14.1 Accounting for [-SA] Characteristics

Let us first consider how this structure can account for the characteristics of [-SA] clauses given in (143a-d), returning in section 2.14.2 to the question of identification. The first characteristic is that [-SA] clauses never occur in embedded clauses. Under earlier theories of the left-periphery, this result could be derived through the fact that the operator occupies the spec-CP position, and so is in complementary distribution with

complementizers. However, under Rizzi's (1997) articulated left periphery hypothesis, this is no longer tenable. Instead, I propose that the restriction to root clauses is due to the nature of the topic operator. The operator is an anaphoric topic operator, and thus looks to discourse for a topic antecedent. If embedded as a syntactic complement, it does not have direct access to discourse, and so cannot occur in such a configuration. It must therefore be in the specifier of the root. Evidence for this comes from the fact that the operator is optionally null⁵⁹ – a typical characteristic of the root (Rizzi, 1997).

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

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

Rizzi, 1997; p.293

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

⁵⁹ The question of *when* the operator can be overt or null is left open at this point. Presumably this turns on discourse conditions, as Rizzi suggests. A

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

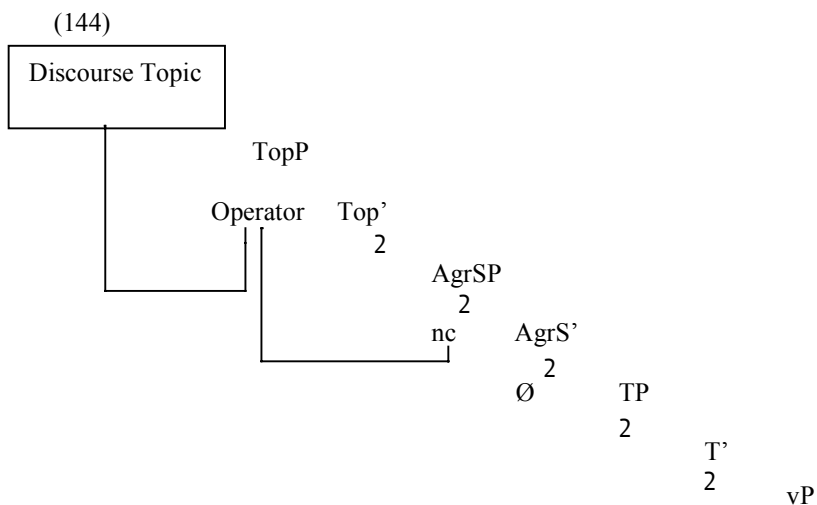
The fourth characteristic is that the subject cannot be quantificational. The anaphoric operator is different from a quantificational operator, in that it 'does not assign a range to its bindee'. Therefore the anaphoric operator cannot be quantificational, and since it is the anaphoric operator that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

2.14.2 Identification in [-SA] Clauses

Let us now consider the identification requirement. In the configuration in (142) above, the null constant has no SA to check its Φ -features, and thus is not identified through this kind of feature checking. The only other possibility is identification through the Operator in spec-TopP position. However, we are now faced with a conflict with Rizzi's claim (from Cardinaletti, 1991) that operators are intrinsically restricted to 3rd person reference. We saw earlier that reference in [-SA] clauses is not restricted to 3rd person subjects, but is free to refer to all persons. How can we resolve this conflict?

clearer understanding of what these conditions are is obviously important, but I must leave it for future study.

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



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

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

In the next section I will present facts about Hebrew null subjects and a theory of discourse identification from Ariel (1990) and Gutman (1999). We will see that Hebrew has person restrictions on null subjects that are accounted for by discourse principles. I will suggest that German null objects are constrained by similar principles. Thus the null operator in German object drop does not have an inherent 3rd person specification. This is important because we see in Swahili [-SA] clauses, the null operator is not restricted to 3rd person. In fact, we will see that the null subject in [-SA] clauses is restricted to 1st and 2nd person, exactly as in the Hebrew case. The overall conclusion that I wish to argue for is that we need not resort to a

stipulation about the nature of the anaphoric operator in order to account for person restrictions in German.

2.14.3 Ariel (1990) and Gutman (1999)

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

- (145) a. *pro* nixshalti ba-mivxan be-historia Hebrew
 failed-1st-sing. In-the-test in-History
 ‘(I) failed the History test.’
- b. *pro* nixshalta ba-mixvan be-Historia
 failed-2nd-sing. In-the-test in-History
 ‘(You) failed the History test.’
- c. **pro* nixshal / nixshela ba-mixvan be-historia
 failed-3rd-m-sg / f-sg in-the-test in-history
 ‘(He/She) failed the History test.’

The Saliency Criterion⁶⁰ includes several ordered pairs, of which (146) shows the more relevant orderings. Thus topics are more salient than non-topics, subjects are more salient than non-subjects, and agents are more salient than non-agents.

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

The second factor in determining accessibility is Unity. Unity refers to the level of syntactic/semantic cohesion that exists between two sentences, e.g., conjoined sentences are less unified than a matrix and embedded clause, adverbs can increase semantic unity, etc. An antecedent that crosses a more unified boundary is more accessible.⁶¹ For example, (147a) is marginal because the antecedent-*pro* relation crosses a sentence boundary that is not

⁶⁰ The Saliency Criterion includes the following ordered pairs, with the element on the left being the more salient.
 1st and 2nd persons > 3rd person
 subject > object > others
 split antecedents interpreted as forming a group > split antecedents not interpreted as forming a group
 matrix antecedents > embedded antecedents
 discourse-topics > non discourse-topics
 antecedents in a Focus Presupposition construction > Antecedents not in a Focus Presupposition Construction.

⁶¹ The Unity Criterion includes the following ordered pairs, with the element on the left being the more unified.
 Embedding > conjoining
 Sentences with parallel time-adverbials > sentences with no parallel time adverbials
 Sentences with consequence adverbials > sentences with no consequence adverbials
 Sentences with other adverbials > sentences with no other adverbials

semantically unified. In (147b), with the addition of semantic adverbials, unity is increased and thus accessibility is increased.

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

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

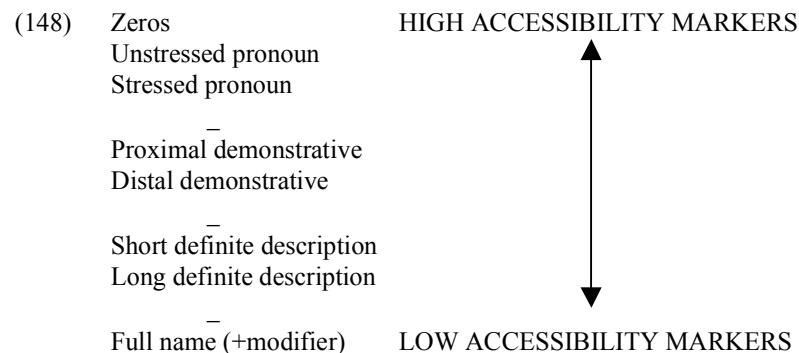
‘Noga just married Shimon, and (they) already got divorced.’

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

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

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

Ariel also assumes that noun phrases differ in the degree to which they depend on antecedents. For example long definite descriptions occur lower on her scale of accessibility than short definite descriptions, which in turn occur lower on the scale than stressed pronouns, etc. At the highest end of the scale are gaps, i.e., null subjects and objects. This is shown in (148). Noun phrases at the higher end of the scale will only recover antecedents with a higher level of accessibility (e.g., topics). Similarly, noun phrases at the lower end of the scale can recover antecedents that are lower in accessibility. Therefore gaps, which are the highest in the scale, require the highest degree of accessibility, and hence are the most restricted.



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

How does this theory account for the Hebrew pattern of subject drop? In past and future tenses, Hebrew allows subject drop of 1st and 2nd person pronouns, but not 3rd person pronouns (although see below). Recall that according to Ariel, null subjects are the highest accessibility markers and thus require an antecedent that is high in accessibility. Ariel claims that 1st and 2nd person antecedents are inherently more salient in the discourse than 3rd person antecedents because they are conversational partners. Because null subjects require the highest level of accessibility, this reduces the frequency of 3rd person null subjects. In fact, Gutman (1999) shows that 3rd person null subjects are not completely unattested, but are considerably less frequent than 1st or 2nd person null subjects. Gutman

shows that when a sufficient level of accessibility is created (through increased saliency and unity), Hebrew allows null 3rd person subjects, as shown in (149).

- (149) Joan_i soxaxa ita axshav be-ivrit, af ki
 Joan chatted-f with-her now in-Hebrew, even though
- pro*_i hevina rak xelek min ha-dvarim she-ha-yalda amra.
 understood-f only part from the-things that-the-girl said-f
- ‘Joan_i was chatting with her in Hebrew now, even though (she_i) understood only part of what the girl said.’

Example (149) is taken from a novel, and demonstrates that in an adjunct clause (high in unity) with a matrix subject antecedent (high in salience), pro-drop is possible in the 3rd person. Gutman argues that this is because the antecedent accessibility is extremely high, as well as the fact that this occurs in literate Hebrew. She argues that the literate medium increases macro (or global) accessibility, making pro-drop much easier.⁶²

In addition to this person restriction in past/future tenses, Hebrew disallows null subjects entirely in the present tense. Gutman (1999) argues this is because of an additional condition that impacts null subjects: null subjects must be syntactically identified in order for discourse identification to be possible. Hebrew present tense is unmarked for person features, and

so null subjects are not *syntactically* identified. This renders discourse identification irrelevant. In the past/future cases, however, both syntactic as well as discourse identification affect the occurrence of null subjects.

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

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

⁶² She argues that literate contexts inherently increase saliency, since Ariel’s original definition of salience was based on processing capacity. Ariel argued that the less salient the antecedent the more taxing it is to link to a null argument. In literature, recovery of identity is considerably easier because of the written medium. In fact, even in English in certain literate contexts of extremely high salience, null subjects are grammatical and very usual. For example, “contains 100% fruit juice” found on a product label. Gutman argues that such examples are licit in English because of the extremely high salience of the antecedent – the product on which the label occurs.

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

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

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

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

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

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

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

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

Unclear = 4

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

null elements, as described by Gutman (1999) and Ariel (1990) for Hebrew, not an inherent restriction on the operator⁶³.

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

2.15 Conclusion

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

⁶³ Intuitions about 3rd person null subject [-SA] clauses vary from speaker to speaker. Of the three native Nairobi Swahili speakers that I have consulted (myself included), the consensus is that 3rd person is "confusing", i.e., there is a strong preference to interpret a [-SA] clauses as 1st or 2nd, and forcing a 3rd person interpretation through rich context conflicts with this preference. This is entirely in-line with the proposal in the text that discourse restrictions prevent 3rd person null subject [-SA] clauses.

identified. Its reference is not restricted if the operator is overt, but when null, there is a preference for more salient antecedents.

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

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

3.0 Swahili data

In this section we will discuss the methodology used to collect the Swahili data, the children who participated, the materials used, the transcription method and the coding procedures.

3.1 Children

The data for this study come from four Swahili speaking children who were recorded in naturalistic settings in Nairobi, Kenya. All sessions were audio recorded using a Sony ProWalkman with an external microphone either by me or by the parent. The children were of varying ages (see Table 3.1), and came from similar socio-economic, ethnic and linguistic backgrounds (see below).

Table 3.1 Age ranges and number of recordings for each child

Child	Age Range	Number of recordings
Hawa (girl)	2;2.01 – 2;6.05	07
Mustafa (boy)	2;0.16 – 2;10.10	23
Fauzia (girl)	1;8.19 – 2;2.07	10
Hassan (boy)	2;10.13 – 2;11.25	04

Hawa’s family lived in a Kikuyu suburb area of Nairobi called Kabete. She lived in a communal environment in which she had lots of play partners and care-givers. She had one younger sister (age 0;6 at the time of her first recording), and an older male cousin (age = 14), who was her primary interactant on several of the recordings. Due to personal problems in the family she was forced to go up-country four months into the study. While her ethnic background is Kikuyu, her primary care-givers considered

Swahili their first language, and so reported that the child was spoken to primarily in Swahili. She could not understand Kikuyu⁶⁴, and knew only a few words of English.

Mustafa was an abandoned child, found by his now-father in an overgrown field outside the family home in the Komarock area of Nairobi. He was estimated to be 4 weeks of age when he was found, and was adopted by the family. The father was a Meru, and the mother was from a neighboring tribe, leaving the couple with Swahili as their common language. For this reason, there was very little influence from any outside languages except English. The amount of English used in the household was minimal, with a few common lexical items being repeatedly used. Mustafa’s primary interactants were his father, mother and step-brother (age = 16). Also, about six months into the study, the family took in a number of homeless teenage girls, and they became Mustafa’s companions.

Fauzia and Hassan were children of two neighboring families in the Majengo slum (also known as Pumwani) of Nairobi’s east side. They spent their time with each other usually, along with several other children of varying ages. There were too many people in their immediate circle to count, but for the purpose of the recordings, their primary interactants were

⁶⁴ This was evident by the fact that she would not respond to people who spoke to her in Kikuyu, and ignored commands in either Kikuyu or English, but complied readily to commands in Swahili. Kikuyu was certainly present in her environment but was not a language of communication or interaction for her.

their parents as well as their common uncle. In this area of Nairobi, the only language spoken was Swahili, with some Sheng (a local street dialect) and a little English. The families were ethnically mixed, and so there is no tribal language to speak of, although there was some Luo and Kikuyu spoken around the neighborhood. All four children in this study were of Muslim parents, and so there was a little Qur'anic Arabic being recited here and there. All four children spoke Swahili of the variety described in chapter 2. It should be noted that Standard Kiswahili (Kiswahili Sanifu) was less common in these areas than English, and in some cases Kikuyu. The only source of Kiswahili Sanifu was from radio broadcasts. A major issue in research in these sorts of complex sociolinguistic communities is language homogeneity. In other words, we must ascertain whether the children in this study all spoke the same dialect of Swahili, and if not, what the differences are. The only way to do this is to examine the language of the adults with whom the children spend most of their time. Thus the speech of adults in the recordings was coded and analyzed. The results show that the adults are remarkably similar in a number of measures (see Appendix 3A for the comparisons and statistics). We conclude that the four children in this study spoke varieties of Swahili that did not differ significantly.

3.2 Data Collection

Each recording session was between 60 and 90 minutes long. At each session, either a parent/care-giver or I (or both) were present to interact with the children. As few other people as possible were included in the conversations, although at times it was difficult to limit neighborhood children from participating. I brought toys to play with, books to talk about

and other conversation pieces, and the sessions were usually conducted as free play time. I would at times try and guide the child to talk about actions and events (and hence use verbs), and sometimes towards using negation, but no formal experiments were conducted.

Transcription was done as soon after the recording as possible. Field notes from the recording sessions and follow-up questions with the parents were used in cases of unintelligible speech or when the reference of a particular utterance was unclear. Thus the number of utterances coded as unintelligible was kept to a minimum. I did all of the transcription myself, with another native speaker of Nairobi Swahili checking segments of the transcripts for accuracy⁶⁵. Furthermore, sections that were unclear or problematic were checked by the both the additional native speaker as well as the parents of the children who were present during the recordings.

Transcription was done in CHAT format (MacWhinney, 2000). Transcription was done with particular attention to the production of verbal and nominal affixes. Transcription was in loose phonetic form, but when pronunciation was drastically different from adult speech, a full phonetic transcription was added in a phonetic tier (%pho). Each speaker was assigned a three letter code and their utterances were transcribed on separate lines. Additionally, a coding tier was used to code children's and parents' speech. The coding was performed as transcription occurred, and so there

⁶⁵ Due to financial limitations, I was not able to hire someone else to check the transcripts. My consultant was good enough to check the transcripts for me, but due to time limitations on his part, he was not able to check all the transcripts fully.

was no bias in the coding procedure⁶⁶. Each child morpheme was coded for function and the number of syllables per word. Additionally, each verbal utterance was coded for transitivity (intransitive, transitive, ditransitive) and for the intended meaning. Intended meaning was determined on the basis of surrounding context, including the child’s previous and following utterances, the adults’ previous and following utterances, and field notes from the recording session. If the intended meaning was unclear, and if the parents who were in that recording session could not elucidate the meaning, then the utterance was coded as unclear and was not included in analyses that make use of intended meaning. Below is an example sentence with the accompanying code tier, and a full gloss and translation in (1b):

- (1) a. FAU: mi namwona hii macho Fau04, line
618
%csc: Spro Ø T pr OA V IND dem N-5 1syll 3syll 1syll 2syll og
ALI: unamwona macho ?
- b. mi na – mw – on – a hii ma–cho
I Ø pres.–OA_{3s}–see–IND these 5–eye
‘I see these eyes’

The child utterance in this example is marked with the three letter code FAU identifying the child. The child’s utterance is coded on the tier labeled %csc (Child Speech Code). See the table of codes in Appendix 3B for explanation of all the codes. Below is an explanation of the codes in this particular example:

Table 3.2 Codes used in example (1)

Code	Meaning
Spro	Subject Pronoun
Ø	Omitted prefix
T	Tense Prefix
Pr	Present tense
OA	Object Agreement
V	Verb root
IND	Indicative Mood final vowel
Dem	Demonstrative
N	Noun root
5	Class marker (in this case noun class 5)
1syll, etc	Monosyllabic word, etc.
Og	On-going present interpretation

In (1a), the child produces a [-SA] clause: the child omits SA in an indicative, matrix clause (see chapter 2 for contexts in which this occurs in adult Swahili). The child uses a reduced subject pronoun (the reduction is not represented in the coding scheme), omits SA, uses a present tense marker and a third person singular object agreement marker. The child uses an indicative mood vowel, and then a demonstrative and noun of class 5 as the object. The syllabic codes indicate the number of syllables in each word in this utterance in sequential order, and ‘og’ indicates that the intention of the child was to convey an ongoing, present tense meaning.

Every utterance produced by all the children was coded in this manner, and the parental utterances in some of the earlier files were coded in this way (see Appendices 3A and 3B). All calculations and statistics were done using CLAN programs (MacWhinney, 2000), making use of these codes. Random analyses were checked by hand to ensure accuracy.

⁶⁶ In some cases, when the function of a particular prefix is unclear or when the intended meaning is ambiguous, it is easy to be biased in assigning codes if a particular analysis is preferred. However, the fact that coding was performed before any analysis was formulated ensured that there was no bias involved.

3.3 Linguistic Measures

The children were of varying ages and at different stages in linguistic development. One measure of grammatical development that has been used extensively is Mean Length of Utterance (MLU). MLU can be calculated on the basis of words/utterance or morphemes/utterance. In Swahili, most utterances consist of a single verbal complex, and so words/utterance is a meaningless measure. I therefore used morphemes to calculate MLU. The specific procedure is outlined below:

- (2) Procedure for calculating MLU
- Start on the second page of the transcript
 - Identify the first 100 non-imitative, non-repetitive, non-formulaic utterances
 - Count each morpheme in those 100 utterances
 - Divide this by 100
 - Morphemes that were counted include the inflectional prefixes (SA, T, OA), verb roots, mood final vowel, infinitive prefix, Grammatical function changing suffixes (applicative, stative, passive, causative, etc.), noun roots, noun class prefixes, adjectives, adverbs, affirmative answers (*ndio*, 'yes', as well as *ehh*, mm-hmm, etc.) negative answers (*hapana*, 'no', as well as ah-ah, etc.), demonstratives, locative suffix, complementizers, copulas, negative prefixes and suffixes, possessives, quantifiers, etc.
 - Not included in the count were repetitions, imitations, formulaic utterances, and any other non-communicative speech.

Table 3.3 below shows the MLUs of the four children at the beginning and end of their recording periods.

Table 3.3 Starting and Ending MLUs for all four children

Child	Starting MLU	Ending MLU
Hawa	1.54	2.46
Mustafa	1.52	3.57
Fauzia	2.97	3.35
Hassan	3.15	4.23

MLU is an effective (but gross) measure of linguistic maturity provided it is used to compare children within the same language. Children acquiring languages which are richly inflected such as Italian (and Swahili) have much higher MLUs (and thus tend to look much more mature) than children of comparable ages acquiring languages which are poor in inflectional affixes, e.g., English and Chinese. For this reason, Valian (1991) suggests using the ratio of verbs to total utterances (I call this measure the V ratio) as an additional measure of grammatical development for cross-linguistic comparison. V Ratio is calculated by dividing the total number of verbal utterances (excluding repetitions and imitations) by the total number of utterances⁶⁷. Table 3.4 shows the V ratios for the four children.

Table 3.4 Starting and Ending V ratios for the four children

Child	Starting V ratio	Ending V ratio
Hawa	.07	.14
Mustafa	.05	.17
Fauzia	.20	.36
Hassan	.30	.40

3.4 Staging Criteria

The database contains data taken from four children of different ages. Furthermore, each child is recorded for different lengths of time. So for example, Mustafa's files span 11 months while Hassan's files span 2 months. For this reason, a staging or ranking criterion is needed in order to meaningfully compare data from all four children. I wish to stress that the results discussed in subsequent chapters are based on both the pooled data described in this chapter, as well as on individual children. I formulated conclusions on the basis of the pooled data because it is easier to observe generalizations using the pooled data, but I always verified my conclusions by comparing the pooled data to the individual children's data. For example, I claim in chapter 4 that bare stems are more frequent in stage 1 than in stage 4. I verify this by looking at the data in the relevant files from each child. The individual subject data are presented in appendix 4C.

I established three semi-independent measures of grammatical maturity and combined them to rank the children relative to each other. These three measures are given in (3):

- (3) I. MLU (in morphemes)
- II. V Ratio
- III. % Mono-syllabic Place holders

We have already discussed MLU and V ratio, and on the basis of those measures we see that the relative ranking of the four children is Hawa<Mustafa<Fauzia<Hassan, i.e., Hawa is the least grammatically

mature while Hassan is the most grammatically mature. One additional measure was used: the proportion of Monosyllabic Place Holders.

Bottari, Cipriani & Chilosi (1993) describe a phenomenon in early Italian that they call Monosyllabic Place Holders (MPH). This is a phenomenon that has been extensively discussed in the child language acquisition literature (see Peters & Menn, 1993, who call them 'filler syllables'; Peters, 2001; Veneziano & Sinclair, 2000, who call them 'additional elements'). Young children use monosyllables at the beginnings of words. These syllables usually occur as reduced vowels from the middle region of phonological space, or nasal consonants. MPHs are found cross-linguistically (e.g., Italian, Bottari *et.al*, 1993; English, Peters, 2001), and are found cross-categorically (i.e., on nouns, verbs, etc.). Bottari *et al.* claim that MPHs are proto-syntactic markers that hold a place in a syntactic representation that the child (for whatever reason) cannot fill phonologically. Crucial for our purposes is the developmental path of MPHs: at the early stages (usually before age 1;8) MPHs are very rare. At some point around age 1;8 there is a spurt in the frequency of MPHs on all categories (verbs, nouns, adjectives and adverbs). For example, Veneziano & Sinclair (2000) in their analysis of one French girl's speech found that the rate of MPHs went from 1.6% at age 1;6.22 to 38.7% at age 1;7.18 (Veneziano & Sinclair, 2000; p.468). After that point, they gradually diminish (see table 3.5 and figure 1).

⁶⁷ An utterance is defined as anything with lexical material in it, including proper names. So a filler utterance such as 'Uhh' would not count, but something like 'Daddy' would.

Table 3.5 Development of MPHs across time

Age	Number of Lexical items	Lexical items preceded by MPH
1;3.2	8	0 (0%)
1;3.16	60	1 (1.7%)
1;4.26	118	2 (1.7%)
1;5.23	154	3 (1.9%)
1;6.22	213	3 (1.4%)
1;7.18	111	43 (38.7%)
1;8.15	190	72 (37.9%)
1;9.3	232	98 (42.2%)
1;10.12	493	103 (20.9%)
2;2.6	190	24 (12.6%)

(Adapted from Veneziano & Sinclair, 2000, p.469, table 1)

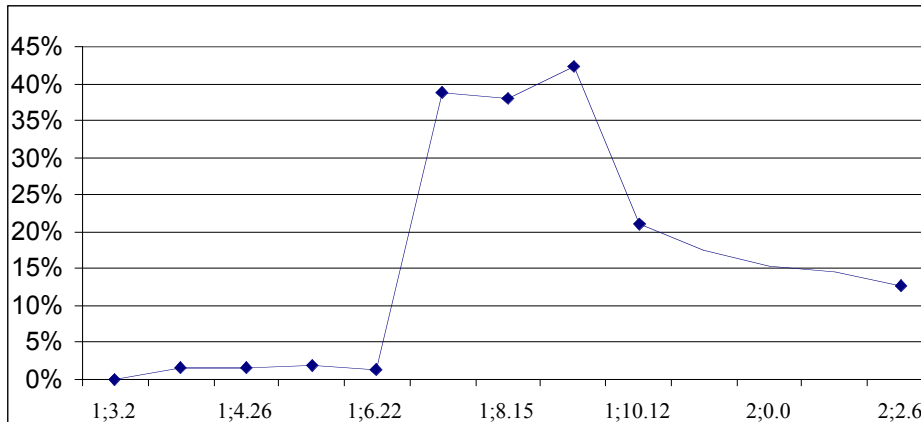


Figure 3.1. Development of MPHs by age
(developed from Veneziano & Sinclair, 2000)

From a peak of approximately 40% they slowly dwindle to 12% at the last data point (See Bottari, *et al.* 1993 and Vollman, 1993, for similar conclusions). This gradual drop-off is characteristic of all children who produce MPHs and thus can be used as a measure of linguistic maturity (see

Peters, 2001 for a review). A high proportion of MPHs indicates that the mapping of morphological forms onto syntactic categories is not yet fully available – the child uses filler syllables as syntactic place holders (according to Bottari *et al.*, 1993). Since MPHs gradually diminish as the child matures we expect that a child who uses fewer MPHs is grammatically more mature than a child who uses larger proportions of MPHs⁶⁸. Swahili children also use MPHs, and I coded these separately from other clearly distinguishable prefixes. As is the case for other languages, Swahili MPHs are usually vowels from the middle area of phonological space (e.g., [↔] or [∅]), or nasals. Of the Swahili tense prefixes, all are CV syllables and hence are clearly distinguishable from MPHs. Similarly, most SA and OA prefixes are CV syllables. However, the following are either single vowels or nasal:

- 2nd singular SA [u]
- 3rd singular SA [a]
- 3rd singular OA [m]

The following procedure was used in order to distinguish MPHs from well-formed Swahili prefixes. If the prefix was fully adult-like in phonological quality (i.e., it was clearly a back rounded [u], a low [a] or a bilabial nasal), it was classified as a well-formed prefix (i.e., not an MPH).

⁶⁸ Of course the limits of this measure are the initial spurt (approximately age 1;8) to the time they fade out entirely, which occurs at approximately age 3;6 (see Vollman, 1999; Peters & Menn, 1993). A further limit to using MPHs is the fact that not all children make use of these filler syllables (Peters, p.c.). So a child that does not use MPHs cannot be assumed to be linguistically mature because a prerequisite for using MPHs as a measure of linguistic maturity is that the child actually has used MPHs in the past. All of the children in this study use MPHs, but to varying degrees, therefore MPHs are an appropriate measure of their linguistic maturity.

If it was reduced or slurred in any manner or if it was unadult-like, it was classified as an MPH. If the prefix was a well-formed prefix but was inconsistent with the contextually-determined meaning, then it was classified as a well-formed prefix used incorrectly (i.e., an agreement error or an incorrect tense). Consider examples in (4a-d). In example (4a) the prefix [u] occurs alone on a verb. In this case the vowel is clearly back and rounded (adult-like), it is not reduced in any manner, and it was clear from context that the referent was 2nd person singular (where [u] is the adult 2nd person singular SA marker). Thus this token was classified as SA. In example (4b), the prefix was well formed (clearly an adult-like [a], which is a 3rd person singular SA marker, or a present tense marker) but did not match the contextually-determined meaning (which was 2st person negative *si*), and so this was classified as an incorrect prefix. If the prefix was a reduced vowel or a non-bilabial nasal (4c and 4d, respectively), because neither of these are adult-like prefixes, the prefix was classified as an MPH.

- (4)
- a. u – tembe – a coded as SA Mus09, line 1408
 target: u – na – tembe – a
 ‘You are walking’
- b. a – tak – i coded as Mus19, line 878
 target: si – tak – i
 ‘I don’t want (it)’
 incorrect prefix
- c. _- tap – a coded as MPH Mus15, line 48
 target: ni – ta – ku – chap – a
 SA_{1s} – fut – OA_{2s} – slap – IND
 ‘I will slap you.’
- d. n – fany – a hivi coded as MPH Haw07, line 1642
 target: a – li – fany – a hivi
 SA_{3s} – past – do – IND this way
 ‘He did it like this.’

Generally MPHs were either [n], [N], [↔], or [∅]. I calculated the proportion of verbal MPHs in the speech of each child. The result of a FREQ count is presented in table 3.6 below.

Table 3.6 Proportion of MPHs on verbs, pooled across files for each child

Child	MPHs	Verbal Utterances	
Hawa	163	474	34%
Mustafa	153	1023	15%
Fauzia	41	653	6%
Hassan	26	512	5%

Notice that the proportion of MPHs is high for Hawa, and somewhat lower for Mustafa, with Fauzia and Hassan showing very few tokens of MPHs. If the proportion of MPHs is an indicator of grammatical maturity (as we propose), then Hassan is the most mature of the four children, with Fauzia, Mustafa and Hawa progressively less mature. Importantly, this is the same order that our MLU and V Ratio calculations suggest. As mentioned earlier, each of these three measures is subject to criticism. However, when a combination of these three measures result in the same order, the ranking procedure gains reliability. I formulated a staging process where each stage was defined in terms of the three measures. The stages were defined according to the following criteria⁶⁹:

⁶⁹ These criteria are arbitrary. The cut-off points were designated so as to create roughly even-sized stages (in terms of number of utterances), but other than that the criteria are not intended to signal anything. They are simply incremental stages in development.

Table 3.7 Criteria for stages

	MLU	V Ratio	% MPHs
Stage 1	<2.5	<.15	>25%
Stage 2	2.5-3.0	.15-.20	15-24%
Stage 3	3.0-3.5	.20-.25	5-14%
Stage 4	>3.5	>.25	<5%

In order to assign children to a stage (or to various stages), I used a point system. I assigned points according to the above schema, so as to get a composite score for each child, and for sections of Mustafa’s development. Table 3.8 shows the scoring system:

Table 3.8 Criteria for assigning scores for the purpose of staging children

	MLU	V Ratio	% MPHs
1 point	2.0-2.5	.01-.15	>25%
2 points	2.5-3.0	.15-.20	15-25%
3 points	3.0-3.5	.20-.25	5-15%
4 points	>3.5	>.25	<5%

The scores obtained for each child on each measure are given in table 3.9, with the composite score being the average of all three scores.

Table 3.9 Composite scores for the purpose of staging children

Child	MLU	V Ratio	% MPHs	Composite Score
Hawa	1	1	1	1
Mustafa	2	1	2	1.7
Fauzia	3	3	3	3
Hassan	4	4	4	4

Table 3.9 shows that the children fall into clear, discrete stages, based on the three language measures. Mustafa is the only child who shows inconsistency, in that the ratio of verbs to total utterances lags slightly behind the other two measures. However, recall that his data span a period of 11 months. If we do a more detailed breakdown of the corpus, we see that Mustafa goes through several stages. Mustafa’s corpus is broken into

three sections, based on the MLUs, V ratio and % MPHs, and the results are presented below.

Table 3.10 Composite scores for sections of Mustafa’s corpus

Age	MLU	V ratio	% MPHs	Composite Score
2;0-2;3	2.24 (1)	.10 (1)	19% (2)	1.3
2;4-2;8	2.67 (2)	.11(2)	18% (2)	2.0
2;9-2;10	3.4 (3)	.20(3)	12%(3)	3.0

We see that Mustafa spans three clear stages. We can therefore say that based on these three measures, Mustafa is comparable to Hawa (who is in stage 1) during ages 2;0-2;3, and is comparable to Fauzia (who is in stage 3) during ages 2;9-2;10. The 5 months in between (2;4-2;8) represent a stage that is between the two. So we have the following breakdown of the corpus by stage:

Table 3.11 Division of the Swahili Corpus by stage

Stage 1	Hawa, Mustafa01
Stage 2	Mustafa02
Stage 3	Mustafa03, Fauzia
Stage 4	Hassan

I have also included in appendix 4C the statistics for each child broken into monthly files for purposes of verification. Thus this staging process is simply a tool that makes the exposition of the facts clearer. In the next chapter we will see the patterns of omission that occur in early Swahili verbal utterances.

Appendix 3A. Adult Similarities

In several places in the text I have made comparisons between children and adults. The adult comparison set comes from the recordings of the children, and thus are comprised mostly of child-directed speech. No other source of spoken Nairobi data is available to me. An implicit assumption that I have made is that the data of the adults forms a homogeneous set, i.e., there are no significant differences from the speech of one adult to the other. I attempt here to justify that assumption.

The decision to make use of adult utterances as a comparison set was made fairly late in the data collection process, and so a full validation of their data is not possible. Ideally, accent ratings could have been conducted with native speaker judges, but there are no independent native speakers available to me. The alternative is a gross, statistical comparison of the composition of the speech of the various adults.

The adults used as the comparison group were the adults who spoke to the children in this study. All the three speakers are native speakers of Nairobi Swahili. All three were born in Nairobi and were raised in Nairobi. All three went to public primary and secondary schools, and none received any university education. They all grew up in low-income, communal areas, as is typical in Nairobi. While not even Nairobi Swahili is a unified dialect, there are significant similarities in the language of the three adults.

I have attempted to quantify these similarities. I calculated the composition of the adults' speech with respect to indicative, subjunctive and

imperative clauses. This is presented below in table 1.⁷⁰ Additionally, I calculated the proportion of [-SA] clauses, [-T] clauses and Bare stems. I then calculated the proportions of applicative morphemes, causative morphemes and passive morphemes, as well as copulas. The correlation coefficients for each data set were calculated (table 2), and they were over 95% for each column.

Table 3A.1 Adult statistics from various files

	Adult from Haw files	Adult from Mus files	Adult from Fau and Has files
Verbal Utterances	1095	1715	640
Indicative	45% (490)	48% (822)	47% (302)
Subjunctive	25% (271)	13% (227)	17% (110)
Imperative	28% (311)	34% (581)	28% (181)
[-SA]	1% (7)	4% (33)	2% (7)
[-T]	0% (1)	1% (8)	0% (1)
Bare Stem	0% (1)	0% (1)	0% (1)
applicative	6% (30)	1% (5)	9% (28)
causative	1% (5)	3% (28)	3% (8)
passive	13% (64)	1% (7)	8% (23)
copula	13% (137)	8% (137)	20% (128)
bare noun	0	0	0

⁷⁰ Percentages for indicative, subjunctive and imperative are out of all verbal utterances. Percentages for all other categories are out of all indicative utterances.

Table 3A.2 Correlation Coefficients for adults

	HAW	MUS	FAU/HAS
HAW	1		
MUS	0.958048	1	
FAU/HAS	0.972979	0.961804	1

The conclusion is that with respect to these salient aspects of Swahili, the adults in this corpus speak a dialect of Swahili that is highly correlated. This certainly is not the ideal measure of dialectal homogeneity, but is an indication that the input to the children is similar.

Appendix 3B. Codes used in transcripts

CHAT format was used in all the transcripts, and CLAN programs were used to perform most of the analyses. Several analyses were checked by hand for verification. Each child's utterance was coded on a separate coding tier labeled %csc (child speech code), and the following list of codes was used. Additionally, the utterances of the adults were coded in the first few files from each child using the same list of codes.

Adj	Adjective
Adv	Adverb
Applic	Applicative suffix
Causative	Causative suffix
Comp	Complementizer, e.g., ya
Cond	Conditional tense, ki
Conj	conjunction
Cont	Continuative tense, ka
Cop	Copula
Dem	demonstrative
Desire	Desiderative reference
Echo	Echo question
Eng	English
Fr	Future reference
fut	Future tense, ta
Hab	Habitual reference
IMIT	Imitation
IMP	Imperative
Inc	Incorrect marking
IND	Indicative
INF	Infinitive
Intentional	Intentional reference, future
Kik	Kikuyu
Lcond	Lexical conditional, kama
Lets	Suggestion reference
LLL	Language other than Kiswahili, English
LNEG	Lexical negation (hapana...)
Loc	Locative

Locsuff	Locative Suffix -ni
NEG	Negative prefix
neg	Negative final vowel
Negative	Negative reference
Negexist	Negative existential, i.e., hakuna
Negpol-xx	Negative polarity item, e.g., yoyote
Num	Number (lexical)
N-xx	Noun-noun class
NA	Na, main verb have, possessive
Ø	Omitted element
OA	Object Agreement clitic
Og	On-going reference
Ogp	On-going reference with a past component (probably present perfect)
Opro	Object pronoun
p	Past tense, li
Past	Past reference
Passive	Passive suffix
Plsuff	Plural suffix e.g., asanteni, or mgonjeni
Poss	Possessive
Possibility	Possibility reference
pp	Past perfect tense, me
Ppneg	Past perfect negation, -ja-
Ppc	Past Perfect Completive, -sha-
pr	Present tense, na
Pref/prefix	Verbal prefix of some sort, MPH
Prep	Preposition
PT	Transcription of the preceding utterance was facilitated by the parent.
qqq	In the speaker tier, nonsensical speech.
Q	Question intonation
Quant	Quantifier, e.g., ngapi
Quote	Used after imperative form of sema
Recip	Reciprocal suffix -an-
Reflexive	Reflexive prefix -ji-
REPET	Repetition
Req	Request reference
SA	Subject Agreement clitic
Should	Permission, request, should I ...? reference
Soc	Social word (e.g., ebu, tafadhali, etc.)
Spro	Subject pronoun

State	Stative present reference
Stative	Stative suffix
SUBJ	Subjunctive Final Vowel
Suff/suffix	Verbal suffix of some sort, ambiguous
T	Tense
Tag	Tag question, like 'eh?'
Top	Topicalization
Ur	Unclear reference
V	Verb
Wh-	Wh- element
xxx	An entire utterance was incomprehensible
Yyy	Part of an utterance was incomprehensible
Zzz	Jibberish, or non-linguistic sounds
^^^	Singing by the child, untranscribed
1, 2, 3syll	Mono-, bi- tri-syllabic word

Appendix 3C. Staging data for each child

The staging process was used for ease of exposition. The data come from four different children who are at different ages. Their linguistic maturity does not necessarily correspond to their chronological age. Furthermore, each child was recorded for a different length of time, making the comparison of data difficult. Thus some sort of staging process was required. It should be mentioned that the staging process does pool data across children, but I believe this does not obscure any facts that are of importance to this study. In other words, the relevant trends are observable in both the staged data as well as the individual children's monthly files. I provide the figures for each child individually in tables 3C.1-4 after the figures, and the monthly files for omissions have been provided in appendix 4C.

The staging data presented in chapters 3-5 suggest that the order of grammatical maturity for these four children is as follows:

- (1) Hawa < Mustafa < Fauzia < Hassan

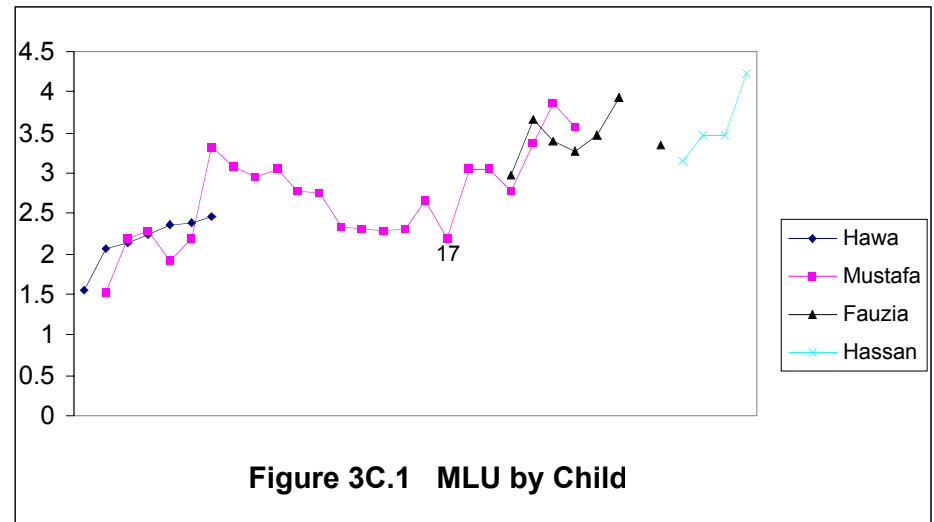
Additionally, Mustafa's files are broken into three sections, ranging from stage 1 to stage 3. The staging criteria were three: MLU (mean length of utterance in morphemes), V Ratio (ratio of verbal utterances to total utterances), and %MPH (the % of ambiguous filler prefixes on verbs).

Below I present the graphs for each criterion individually. I have arranged the data for each child roughly relative to the four stages, with Hawa's data on the left, Mustafa's data spanning three stages, Fauzia's data in stage 3 and Hassan's data on the right in stage 4.

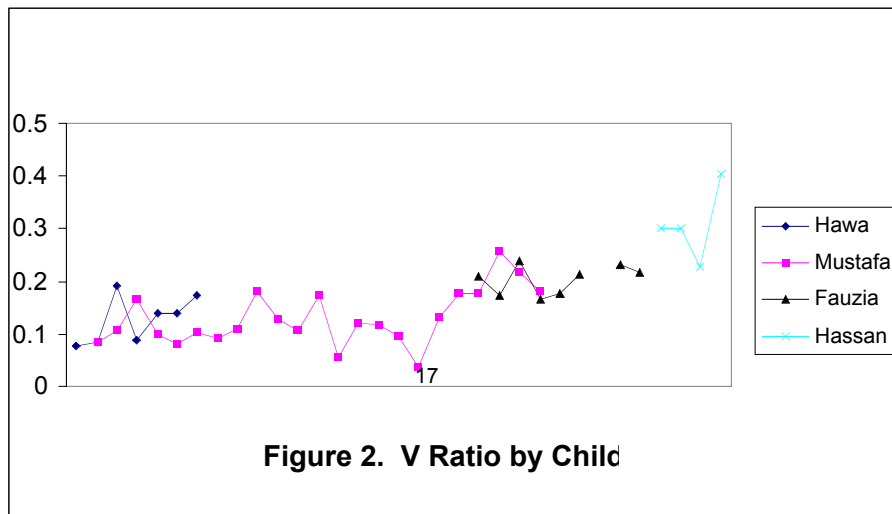
The first figure (figure 3C.1) shows MLU. This is the clearest of the three graphs, showing that the children's data is in fact ordered as in (1).

Notable points in this data are the low in Mustafa's file 17 (noted on the graph). This entire period was a time when Mustafa was suffering from a terrible sickness. He was in and out of hospital, and this data recording was a particularly bad one. We see that this affected more than just his MLU, as his V Ratio and %MPH scores are significantly affected as well (see figures 3C.2 and 3C.3 below, where data point 17 is noted). Mustafa's sickness began at file 12 and ran through file 18. Nevertheless, the overall trend remains clear that in the early files Mustafa was significantly less mature on all three counts than in later stages.

Figure 3C.2 shows the V Ratio for each child, again arranged roughly relative to the stages in which they have been assigned. The vertical axis represents age, and the horizontal axis represents data points (at

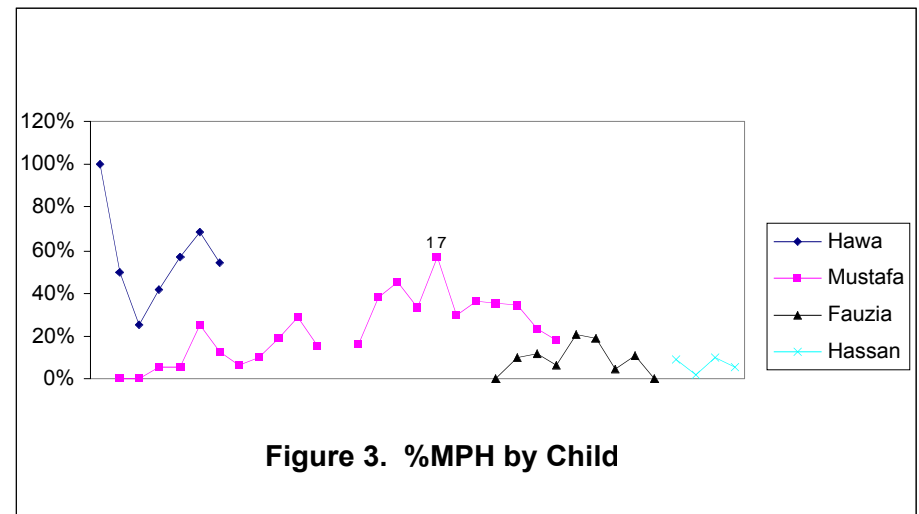


approximately 2 week intervals). There is a considerable amount of variation from file to file, but the overall trend remains clear. Notice data point 17 for Mustafa which is particularly low.



Finally, figure 3C.3 shows the %MPH rates for each child. This figure appear to contradict the general trend of the previous two figures, but recall the developmental pattern of MPHs discussed in chapter 3. We saw a graph from Veneziano & Sinclair (2000) which describes the development of MPHs as occurring on the following path: initially MPHs are rare in the speech of the child, then there is a spurt in MPHs, followed by a gradual decline. Hawa's rate of MPHs is relatively high compared to Mustafs's early files (which are supposed to be in the same stage). However, there is variation from child to child as to how frequent MPHs are overall in their speech. Some children never produce MPHs, going in stead from the single

word stage to the multi-word stage without a period of MPHs (Peters, 2001). Some children produce MPHs at rates of over 50%, others at 20%.



Hawa produces MPHs at a rate of 30%-60% (the first data point has very few utterances, and so the rate of 100% is misleading), while Mustafa produces MPHs at a rate of approximately 20%. The surge in MPHs around data point 17 can be attributed to his sickness. The other two children produce very few MPHs, as would be expected given their other linguistic measures.

Below I present the figures for each child by file. The final file for Fauzia is not included in the staging data at any point because of its minimal size (hence difficulty in calculating MLU and the other two criteria).

Table 3C.1 Age, MLU, V Ratio and %MPH for Hawa

File	Age	MLU	V Ratio	%MPH
HAW01	2;2.01	1.54	0.077364	100%
HAW02	2;3.01	2.06	0.082863	50%
HAW03	2;3.14	2.13	0.190476	25%
HAW04	2;3.27	2.23	0.087028	41%
HAW05	2;4.24	2.37	0.140625	57%
HAW06	2;5.22	2.39	0.141026	68%
HAW07	2;6.05	2.46	0.172468	54%

Table 3C.3 Age, MLU, V Ratio and %MPH for Fauzia

File	Age	MLU	V Ratio	%MPH
FAU01	1;8.01	2.97	0.208108	0%
FAU02	1;9.01	3.66	0.172542	10%
FAU03	1;9.14	3.4	0.238213	12%
FAU04	1;10.02	3.28	0.166144	6%
FAU05	1;11.01	3.47	0.177711	21%
FAU06	1;11.27	3.93	0.213256	19%
FAU07	2;0.26	6.1	0.591195	5%
FAU08	2;1.07	3.35	0.232628	11%
FAU09	2;1.22	3.26	0.217391	0%

Table 3C.2 Age, MLU, V Ratio and %MPH for Mustafa

File	Age	MLU	V Ratio	%MPH
MUS01	2;0.16	1.52	0.085714	0%
MUS02	2;1.05	2.19	0.10625	0%
MUS03	2;2.01	2.29	0.16701	6%
MUS04	2;2.22	1.93	0.098101	6%
MUS05	2;3.05	2.2	0.080425	26%
MUS06	2;3.17	3.31	0.101266	13%
MUS07	2;4.01	3.08	0.09324	6%
MUS08	2;4.22	2.94	0.110048	10%
MUS09	2;4.29	3.04	0.179153	19%
MUS10	2;5.12	2.79	0.127789	29%
MUS11	2;5.20	2.76	0.105932	15%
MUS12	2;6.03	2.33	0.171429	100%
MUS13	2;6.14	2.32	0.056075	17%
MUS14	2;6.24	2.28	0.121739	38%
MUS15	2;7.01	2.3	0.118998	45%
MUS16	2;7.17	2.65	0.094787	33%
MUS17	2;7.30	2.19	0.036885	57%
MUS18	2;8.11	3.04	0.132394	30%
MUS19	2;8.25	3.05	0.175393	36%
MUS20	2;9.06	2.78	0.174731	35%
MUS21	2;9.16	3.38	0.257373	34%
MUS22	2;10.03	3.87	0.217125	23%
MUS23	2;10.10	3.57	0.179739	18%

Table 3C.4 Age, MLU, V Ratio and %MPH for Hassan

File	Age	MLU	V Ratio	%MPH
HAS01	2;10.13	3.15	0.301616	9%
HAS02	2;10.27	3.46	0.301837	2%
HAS03	2;11.11	3.47	0.228228	10%
HAS04	3;0.01	4.23	0.405612	6%

4.1 Introduction

Recall from chapter 1 we saw that children learning a variety of languages allow root bare verbs. We saw that children acquiring Sesotho allow bare verbs at early ages before producing verbs with filler ‘shadow’ syllables, and then going on to adult-like utterances (Demuth, 1992; 1994). Children acquiring Siswati follow the same developmental order (Kunene, 1979). Children acquiring Quechua, an agglutinative language spoken in Peru, use bare verbs very frequently at early stages - 57% of all verbal utterances (Courtney, 1998). We also saw that children acquiring Inuktitut produce bare verbs, which are ungrammatical in adult Inuktitut (Swift & Allen, 2002). In all of these adult languages bare verbs are unattested, and so this constitutes a genuine departure from the adult norm.

We also saw that errors of omission are very frequent in child language (bare verbs being an example of such an error), but errors of commission are extremely rare. The one exception to this generalization is the case of Root Infinitives (RIs). We saw that children learning languages such as German, Dutch, French, etc. allow root clauses with infinitival verbs. However, children learning other languages such as Italian, Spanish, etc. produce neither RIs nor bare verbs. Thus we developed a typology of languages, as summarized in table 1.12 in chapter 1, reproduced here:

Table 1.12 Summary of languages that allow RIs, disallow RIs and allow bare verbs.

True RI languages	Non-RI languages	Bare Verb Languages
German	Italian	English
Dutch	Spanish	Sesotho
French	Catalan	Siswati
Swedish	Japanese	Quechua
Icelandic		Inuktitut
Russian		

How does Swahili fit into the typology described in table 1.12?

We know that Swahili has an infinitive prefix (*ku*), and so RIs in principle should be available to children. However, Swahili may behave like Italian (which also has an infinitive, but which is not an RI language). Finally, Swahili may behave like its closely related language Sesotho in exhibiting bare verbs. The inflectional morphology of Swahili is particularly interesting because it independently and distinctly marks Agr and T. As we will see, this proves crucial in distinguishing between several influential theories in language acquisition. We will investigate the question of whether Swahili is an RI language, a bare verb language or neither, as well as the development of inflectional affixes in the Swahili verbal complex.

This chapter is organized as follows. In section 4.2 I discuss five theories of language acquisition. The first approaches the problem from a phonological perspective and attempts to account for the general tendency towards omission. The remaining four approaches are syntactic in nature. In section 4.3 I consider the predictions that each theory makes for Swahili,

and in 4.4 I introduce the results. In section 4.5 I return to these theories and re-evaluate them in light of the Swahili data. Section 4.6 includes a summary and some concluding remarks.

4.2 Theories of language acquisition

In this section we will discuss five theories of the acquisition of functional elements, each approaching the problem from varying perspectives and drawing on different data. Given that Swahili overtly represents functional material in such a clear and unambiguous manner, these theories are particularly relevant to the acquisition of Swahili. The theories fall into three basic classes. The first considers the omission of functional elements such as determiners to be a result of production constraints. Thus a full syntactic representation is postulated, with production limitations constraining the expression of the entire representation. The second theory considers child Root Infinitives and other functional omissions to be the result of truncation of higher portions of the syntactic representation. The final class of theories, often called Full Clause Theories, argues that early grammars project all the way to the CP in all cases, but that specific deficits in particular heads result in the omission of particular features. After discussing each proposal, I state in section 4.3 the predictions that each theory makes for child Swahili. It is here I believe Swahili may be most enlightening since, as we will see, each theory makes precise predictions about the acquisition of Swahili. In section 4.4. I present the data and return to these theories of acquisition in section 4.5.

4.2.1 Metrical Omission Model (Gerken 1991; Gerken & McIntosh, 1993)

It has long been noted that children learning English typically omit determiners in obligatory contexts (Brown, 1973; Radford, 1990; Hoekstra & Hyams, 1995; Hoekstra, Hyams & Becker, 1996).

- | | | | |
|-----|----|------------------|-----------------|
| (1) | a. | Eve sit floor | (Brown, 1973) |
| | b. | Hayley draw boat | (Radford, 1996) |

Gerken (1991) and Gerken & McIntosh (1993) propose that children have less processing capacity than adults, which leads to constraints on the production of phonological segments. These production constraints prevent children from expressing a full linguistic representation by forcing them to produce only syllables that conform to their language-appropriate metrical pattern. Specifically, children omit weakly stressed syllables that either fall outside the metrical pattern that is typical of their language or that occur as part of a non-typical metrical foot. For example, English has a predominance of trochaic feet, and so weak syllables that occur in pre-trochaic position or weak syllables that occur as part of an iambic foot are subject to omission (see 3 below). Gerken (1991) discusses the nature of these production limitations. She suggests that previous production accounts (e.g., L.Bloom, 1970; P.Bloom, 1989; Valian, 1989) suffer from the limitation that they must posit a system of meta-awareness in order for the child to know when to omit material. Most production-limitation accounts argue that sentence complexity adds to processing load, and thus the child omits material in complex sentences. Gerken quite rightly points out that this means that the child must have a system of meta-awareness that will tell the child when something is overly complex. This

system must also be able to pragmatically identify material that can be omitted (i.e., redundant material, old information, etc.). These additional systems presumably increase processing complexity, and thus are not desirable in a processing account of omission.

Instead, Gerken proposes that children's omissions are the result of mechanisms within the production system itself. She proposes a system of production that makes use of rules and templates. The original linguistic message goes through several ordered levels of rules and templates (semantic > syntactic > morphological > foot formation > head location > phonological rules and templates) before the utterance is produced. She proposes several principles that account for children's omissions. First, the child has limited processing resources. Second, templates draw less resources than rules do. And third, using resources earlier in the process (i.e., closer to the message as opposed to the produced utterance) is less taxing. Thus making use of templates at the semantic level is less taxing than making use of templates at the syntactic level.

There are two possible metrical patterns in languages, and individual languages differ as to their preferred metrical pattern. A trochaic foot is one in which the syllabic stress pattern is Strong-Weak (2a), while an Iambic foot has the pattern Weak-Strong (2b)⁷¹. The trochaic foot is the basic meter used in the production of English words, while French makes use of Iambic feet as the basic metrical pattern.

⁷¹ Gerken argues (following Hayes, 1982; Kelly, 1988) that binary feet are preferred over both ternary feet and unbounded feet (both of which are possibilities in adult languages). She thus only considers the simplest possibilities of trochaic and iambic feet. I follow her on this. Furthermore, the strength of a syllable is determined by the pitch, the stress, the length, etc. (Gerken, 1991; Gerken & McIntosh, 1993).

- | | | | |
|-----|----|---------|---------------|
| (2) | a. | monkey | Trochaic foot |
| | | S - W | |
| | b. | giraffe | Iambic foot |
| | | W - S | |

Gerken proposes that there is a trochaic template operating at the level of foot formation or head location. She claims that this may be because trochaic feet are easier to plan and produce (in terms of motor skills) than iambic feet. In fact, Allen & Hawkins, (1980) were the first to suggest this idea, and several others have argued in favor of such a position (e.g., Fletcher, 1985; Echols & Newport, 1992; Fikkert, 1992; Archibald, 1995; La Belle, 1999). Gerken points out, however, that if this were the case, it would put languages in which iambic feet predominate (e.g., French) at a serious disadvantage, and so she suggests that the trochaic template may develop from the English-speaking child noticing that trochaic feet are more common in her language.

Because of the constraints on production, children must target material for optional omission. The Metrical Omission Model (MOM) states that weakly stressed syllables are targeted for omission, never strong syllables. Furthermore, children prefer binary feet, and cannot leave any strong syllables unparsed⁷². Because English is a trochaic language, English speaking children apply a trochaic template to their production system and maximize the number of trochaic feet in any particular string. Furthermore, in a trochaic language such as English, children tend to drop weak syllables that are part of an iambic foot (the dispreferred pattern, as in (3a)), and weak syllables that precede a fully formed trochaic foot (as in 3b).

⁷² Weak syllables, however, may be left unparsed.

Crucially, however, weak syllables in a trochaic foot are not subject to omission (3c).

- (3) a. the ball
[W-S] Weak Syllable in Iambic foot
↑ Subject to Omission
- b. the mon key
W [S - W] Weak pretonic syllable
↑ Subject to Omission
- c. mon key
[S - W] Weak syllable in Trochaic foot
↑ Not Subject to Omission

Notice that in both (3a) and (3b) the syllables that are subject to omission are determiners. This has the desirable prediction that determiners will be particularly prone to omission because of the metrical pattern of English and the placement of determiners in English (cf. examples (1) in section 1.2.1). As the child's processing capacity improves, the omission of determiners decreases until they reach adult-like proportions.

Demuth (1994) uses the MOM to show that Sesotho children omit noun class prefixes based largely on these principles. Sesotho noun class prefixes (like Swahili noun class prefixes) are monosyllabic (usually CV) prefixes that attach to either a mono- or di-syllabic noun stem. Omission of prefixes occurs on nouns as well as agreement prefixes on other elements such as demonstratives:

- (4) [kolo] [sá-ne] Sesotho
adult form: se-kólo sá-ne (from Demuth, 1994, p.129)
'that school'

According to Demuth, in (4) the agreement prefix *se-* on the demonstrative *kólo* is omitted by the child because it occurs as a pretrochaic weak syllable. Demuth extends this analysis to account for why Sesotho children omit preverbal inflectional affixes. She shows that Sesotho has penultimate lengthening which produces trochaic feet at the end of all prosodic words. This creates a trochaic bias, which when coupled with the principles of the MOM predict omission of preverbal affixes. Following are her examples (18) from a child aged 2;1:

- (5) a. ta hâ:na
adult form: [ke - a - hán - a +]
SA_{1s} - pres - refuse - IND
'I refuse.'
- b. áy shépa
adult form: [ó - a - sháp - a +]
SA_{1s} - pres - lash - IND
'S/he is lashing.'

In (5a), the preverbal affixes have been reduced from *ke-* to *ta-*, while in (5b) the preverbal affixes are reduced from *ó-a* to *áy*. Demuth claims that this reduction is because of their position as weak pretonic syllables. She further suggests that children operate under a Minimal Word Constraint which she states as in (6):

- (6) A prosodic word contains a foot

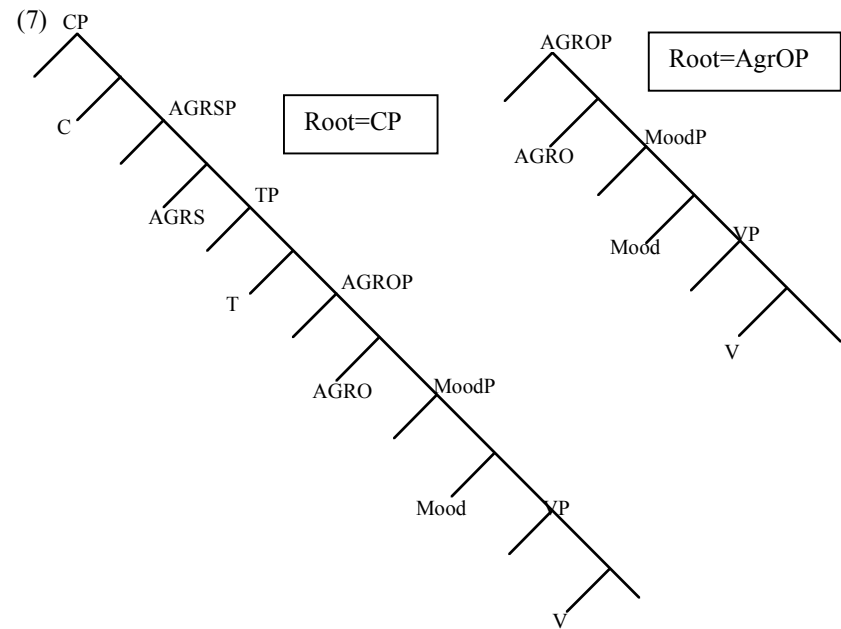
Thus the examples in (5) contain a foot, and satisfy the Minimal Word Constraint. However, Demuth claims that 'syllables falling outside of the foot will be treated as extrametrical and subject to reduction or deletion' (p.128). Thus the prefixes in (5) are reduced because they fall outside the basic trochaic foot.

These results being from Sesotho, a related Bantu language, may be indicative of what we may expect in Swahili. We will discuss the predictions MOM has for Swahili in section 4.3.1.

4.2.2 Truncation (Rizzi 1994)

Rizzi (1994) proposes that early grammars optionally fail to project the entire CP, producing truncated structures. Rizzi's primary concern involves RIs (see section 1.2.3). Recall that RIs are optional, they occur in unraised positions, and typically do not occur in *wh*- questions. Based on these facts, Rizzi proposes a structural account of RIs. Rizzi proposes that young children differ from adults in that the adult axiom of 'Root=CP' is not operative in child grammar. Children can optionally specify the root as any projection, and can thus have a truncated structure. While the adult must project a structure as high as the root CP for every declarative clause, the child can optionally project to a lower position, e.g. AgrOP (see 7). In the case of RIs, the child projects to a position lower than TP, and hence produces a tenseless clause.

The optionality of RIs is accounted for by the absence of the axiom root=CP. Since children can optionally specify the root as any projection in the structure, RIs can optionally occur. Furthermore, according to Rizzi, RIs occur in structures that are truncated below TP. Thus RIs occur in structures that are missing all projection above TP, including the CP. Recall that RIs rarely occur in *wh*- questions: a fact that is elegantly accounted for under this proposal since the presence of a *wh*-question necessarily entails a CP projection, hence everything below the CP (including Tense) must also be projected.



A crucial property of truncation is that when a particular projection is specified as the root by the child, all the structure up to that point is projected. For example, if the child projects up to the CP (such as when a *wh*- question occurs), it is not possible to leave out any intervening projections, such as TP. Hence the absence of RIs in *wh*- contexts. This means that any omission in the grammar of the child occurs at the higher periphery, and cannot target specific projections internal to the sentence.

The remaining three theories assume that children project each sentence all the way to CP, and so are often referred to as Full Clause Hypotheses. However, in order to account for the particular unadult-like characteristics of child speech, specific heads are assumed to be

underspecified⁷³. Thus the difference between adults and children is localized to a particular head or heads. I will now discuss the three versions of underspecification that I believe have the most relevance to Swahili.

4.2.3 Underspecification of Tense (Wexler 1994)

Wexler (1994) points out that infinitives in adult languages like German and Dutch occur in embedded clauses, they license PRO, they can be used gerundively, etc.

- (8) a. John tried [PRO to eat all the plums]
 b. PRO to run everyday is good for your health

These properties are thought to derive from the absence of tense in infinitives. Wexler argues that children have the option to underspecify tense in main clauses, giving rise to RIs. This leads to the natural correlation of null subjects and RIs discussed in chapter 1, as PRO is licensed in non-finite contexts.

There are three pieces of evidence that suggest an underspecification of Tense in child RIs:

- (i) The form that occurs is an infinitive, which in the adult language is tenseless;
- (ii) RIs occur in unraised positions, which correspond to tenseless verbs in the adult language;
- (iii) RIs correlate with null subjects, which by hypothesis are PRO licensed in tenseless environments.

⁷³ Underspecification means that a feature matrix on a particular head is empty. Often the feature matrix is an entire feature bundle (e.g., $\bar{_}$ -features), but sometimes it may be a single feature (e.g., a case feature). Alternatively, this could mean that the feature has no value.

Furthermore, Wexler argues that English bare verbs as in the examples in (9) and RIs in German, French, etc. result from the same underlying mechanism.

- (9)
- | | | | |
|----|--------------------------|----------------|---------------|
| a. | He fall down | Nina (2;1.29), | Suppes, 1971 |
| b. | Hayley draw boat | Hayley (1;8), | Radford, 1996 |
| c. | He bite me | Sarah (2;9) | Brown, 1973 |
| d. | It only write on the pad | Eve (2;0) | Brown, 1973 |

Wexler groups English bare verbs, bare participles / null auxiliaries and null copulas with RIs and refers to them as Optional Infinitives, arguing that they are all the result of the underspecification of Tense. Unlike Rizzi's Truncation Hypothesis, Wexler proposes that the difference between child grammar and adult grammar is specific and minimal rather than global in nature. The nature of the difference lies in the particular nature of an individual head - Tense.

4.2.4 Underspecification of Agr (Clahsen *et al.* 1996)

Clahsen *et al.* are concerned with the correlation between the acquisition of lexical knowledge (in terms of inflectional morphology) and certain syntactic effects. Putting aside the question of lexical learning versus full clauses⁷⁴, let us focus on their underspecification proposal. Clahsen *et al.* are concerned with accounting for four pieces of evidence that are established in Clahsen (1990) for child German, listed in (10):

⁷⁴ They are primarily concerned with the Lexical Learning Hypothesis which states that the acquisition of inflectional morphology drives the development of syntax in general. Thus they investigate the acquisition of subject-verb morphology and try to draw parallels in real time with the development of other syntactically related phenomena. This is not crucial to what follows, and so I shall put it aside.

- (10)
- a. Subject-verb agreement, accusative case and dative case have not been acquired.
 - b. Finite verbs raise to first or second position, while non-finite verbs remain in clause-final position.
 - c. Subjects never intervene between the finite verb and negation (hence, in contrast to adult German, subjects never raise).
 - d. wh-elements and complementizers are completely unattested.

They propose that children have a single functional projection above VP which is specified for +finite. This position cannot be identified with IP or AgrSP because the specifier of this position is not restricted to subjects. This position cannot be identified with CP because at this stage German children never produce wh- elements or complementizers (Clahsen, 1990). Therefore Clahsen *et al.* refer to this projection as FP – a general functional projection that is specified as <+finite>. They describe the feature <+finite> as a syntactic feature that allows its head to function as the landing site for a finite verb. The only reason they posit such a projection is because finite verbs quite robustly raise to first or second position. So FP accounts for the position facts in early German (finite verbs raise while RIs remain in final position)⁷⁵. The absence of any other features and structure also accounts for the absence of wh- questions and complementizers. They go on to show that as the subject-verb agreement paradigm is acquired, syntactic processes associated with AgrSP develop, i.e., subject raising to [spec, AgrSP].

⁷⁵ They show that a small proportion of finite verbs remain in final position (approximately 12% of all finite verbs), but almost no nonfinite verbs ever raise. They account for this by adding the specification of +V to the lower verbal position, making it compatible with both finite and nonfinite verbs. Therefore only finite verbs are allowed to raise but all verbs may remain low. This is not relevant to our discussion, so I will not discuss it further.

I interpret Clahsen *et al.*'s FP position as most closely resembling TP. According to Clahsen *et al.*, the FP projection carries <+finite> features⁷⁶, which can include temporal features, as well as agreement, aspectual and mood features (Rizzi, 1997). However, since subject-verb agreement is seen as being integral to AgrSP (which has not developed, according to Clahsen *et al.*), finiteness here cannot include agreement features. Aspect and mood are generally seen not to affect verb raising in German, and so it is unlikely that FP includes aspectual or mood specification. We can therefore interpret finiteness in this case as being tense specification. Thus, another way of interpreting Clahsen *et al.*'s claim is that AgrSP (as well as AspectP and MoodP) is underspecified in early

⁷⁶ They describe this feature as being hosted by FP, and allowing the hosting head to act as the landing site for finite verbs. However, 'finiteness' is a general descriptive term that is usually composed of several distinct features. For example, Rizzi (1997) describes finiteness as including tense, aspect, mood and sometimes agreement features. I assume this definition of finiteness. What is important for Clahsen *et al.* is that there be one and only one position to which verbs and only verbs may raise. Therefore this position cannot be associated with subjects (hence AgrSP is ruled out), and it cannot be associated with wh- elements or complementizers (hence CP is ruled out). This is why I say that their FP can be thought of as TP.

Table 4.1 Summary of possible INFL features for ATOM

	Verb form	Subject	Examples
Tense = present, +agreement	-s	NOM	He cries
Tense = present, -agreement	OI	ACC	Him cry
Tense = past, +agreement	-ed	NOM	He cried
Tense = past, -agreement	-ed	ACC	Him cried
-Tense, +agreement	OI	NOM	He cry
-Tense, -agreement	OI	GEN?	His cry

(Table taken from Schütze, 1997, p.232)

Schütze & Wexler’s argument holds appeal for two reasons. First, it draws attention to an additional syntactic effect of underspecification: Case. Thus case effects are now added to the list of syntactic effects that correlate with OIs: null subjects, no verb raising, and absence in wh-questions. This adds to the evidence that RIs are not speech errors or the result of production problems, but a true syntactic phenomenon. Second, they provide a theory of underspecification that is as precise as the other theories, but that allows for a greater amount of underspecification. Thus ATOM has an increased descriptive power because it allows more than a single underspecified head, but less predictive power because it introduces an additional degree of freedom into the model.

Swahili, as we saw in the previous chapter, marks tense and subject agreement as independent bound prefixes, making Swahili an ideal language to test these theories on. This is especially true for the underspecification theories since they each make claims about an inflectional head that is marked in Swahili as a unique and distinct prefix: SA and T. In the next section I will discuss the specific predictions that each theory makes for Swahili.

4.3 Predictions for Swahili

None of the theories reviewed in the previous section were intended to account for Swahili. Thus we do not intend this to be a criticism of the theories. However, as generalizability is a desirable property of any theory, Swahili allows us to test how generalizable each theory is. Furthermore, each of the theories discussed in the previous section make precise predictions about the kinds of things Swahili children should allow.

4.3.1 Metrical Omission Model

The Metrical Omission Model is important for us to consider for two reasons. First, Swahili has a trochaic bias, as we saw in section 2.2. Furthermore because the inflectional morphology is primarily prefixed, these prefixes are potentially in the contexts that Gerken identifies as prone to omission. We will discuss this in more detail below. The second reason that this model is important for us to consider is because Demuth (1994) uses it to account for Sesotho noun class omission and Sesotho verbal prefix omission/reduction. She argues that because Sesotho is a language that has penultimate lengthening and a bias for trochaic feet, verbal and noun class prefixes are usually either in pre-trochaic position or weak syllables in iambic feet. Thus verbal and noun class prefixes should be subject to omission/reduction by the Metrical Omission Model. Because Swahili is also a trochaic language, we expect the same pattern of omission as we see in English and Sesotho:

- Omission of pretonic weak syllables, e.g., **W** [S – W]
- Omission of weak syllables in iambic feet, e.g., [**W** – S]
- No omission of weak syllables in trochaic feet, e.g., [S – W]

We can use this as a guide to assign metrical structure to a typical Swahili verbal complex. The minimal verbal complex in Swahili consists

of SA-T-V-IND. Recall that primary stress in the Swahili verbal complex is always on the penultimate syllable, and secondary stress is on SA⁷⁹. Recall also that the strength of a syllable is determined by stress, length, pitch, etc. The majority of verb stems in Swahili (where a verb stem is the root + obligatory final vowel) are disyllabic (Ashton, 1947; Maw & Kelly, 1975; Vitale, 1981; Krifka, 1995), and so primary stress occurs on the first syllable of the verb stem. We saw in section 2.2 that the SA marker carries secondary stress (an argument that a phonological word boundary exists between T and the rest of the verbal complex). Therefore the penultimate syllable in the verbal complex is parsed as strong, as is the SA marker (which carries secondary stress).⁸⁰ Thus the pattern of strong and weak syllables in a typical Swahili verbal complex is given in (13). In this example I show the stress pattern on the first line, followed by a syllabic break-down of the sentence. This is followed by a morphemic analysis of the sentence and then a morpheme-by-morpheme gloss and finally a translation.

- (13) S W S W
 Syllabic: nì – me – fí – ka
 Morphemic: nì – me – fík – a
 SA_{1s}-p.perf.-arrive-IND
 ‘I have arrived.’

⁷⁹ The only cases in which stress is not on SA is when T is multisyllabic, e.g., -mesha-. These tense markers are relatively rare in child Swahili, with the vast majority of tense markers in child Swahili consisting of monosyllabic markers (primarily *na*, *li*, *me*, and *ta*).

⁸⁰ Note that this differs from the analysis proposed by Demuth (1994) for Sesotho in which Sesotho is not assumed to have secondary stress, but rather penultimate lengthening is the only process applied. As we will see, this has dramatic effects on our predictions for Swahili.

The rules of the MOM state that the parser:

- (14) a. Leaves no strong syllable unparsed,
 b. Maximizes the number of binary feet,
 c. Prefers a trochaic pattern.

Using these rules, the only possible parse for this string in the child production of Swahili is as follows:

- (15) [S – W] [S – W]
 nì – me – fí – ka

The outcome of a metrical analysis of a typical Swahili verbal complex is two trochaic feet. Thus the MOM predicts that in such utterances there should be no omission by children. Let us now consider other more elaborate verbal complexes. We will consider three cases: first, cases in which the verb stem is more than two syllables (examples 16-17); second, cases in which syllabic suffixes (e.g., applicative) occur between the verb root and the mood final vowel (example 18); and third, cases in which OA occurs between T and the verb root (example 19). In these examples I provide a syllabic decomposition of the string as well as a morphemic decomposition, followed by a gloss and a translation.

(16) syllabic morphemic	S W W S W ni – me – an – gu – ka ni – me – anguk – a SA _{1s} -p.perf.- fall – IND ‘I have fallen.’	→ [S – W] W [S – W]	3 syllable V stem
(17) syllabic morphemic	S W W W S W ni – li – ten – ge – ne – za ni – li – tengenez – a SA _{1s} - past – fix – IND ‘I fixed (it)’	→ [S – W] W W [S – W]	4 syllable V stem

(18) syllabic morphemic	S W W S W a - li - pi - ki - a a - li - pik - i - a SA _{3s} -psat-cook-applic-IND 'He cooked for (someone).'	→ [S - W] W [S - W]	Applicative suffix
(19) syllabic morphemic	S W W S W a - li - ni - pi- ga a - li - ni -pig-a SA _{3s} -past-OA _{1s} - hit-IND 'He hit me.'	→ [S - W] W [S - W]	OA prefix

Consider first the cases in which the verb stem is more than 2 syllables long. Example (16) shows a stem composed of three syllables, and (17) shows a stem composed of four syllables. Using the rules of the MOM (14), the only possible metrical analysis is given to the right of each example. Notice that in both cases weak syllables are left unparsed between two trochaic feet. According to the MOM these medial weak syllables are subject to omission. Therefore unlike verbs that have disyllabic stems, longer verb stems do give rise to potential omission through the MOM. However, the syllables that are subject to omission in both cases are onset syllables of the verb stem. Therefore the inflectional prefixes are not subject to omission under the MOM.

Similarly in the case of a syllabic suffix, as in (18), a medial weak syllable between two trochaic feet occurs, which is subject to omission. But in this case, as in the previous two cases, the weak syllable that is subject to omission is the onset syllable of the verb stem, not the inflectional prefixes. And finally the case of OA prefixes (19): the weak medial suffix that is subject to omission is the OA prefix. Therefore according to the MOM, OA should always be subject to omission because it is always a pretonic weak syllable. However, as in all the previous cases, the inflectional suffixes SA and T are never subject to omission.

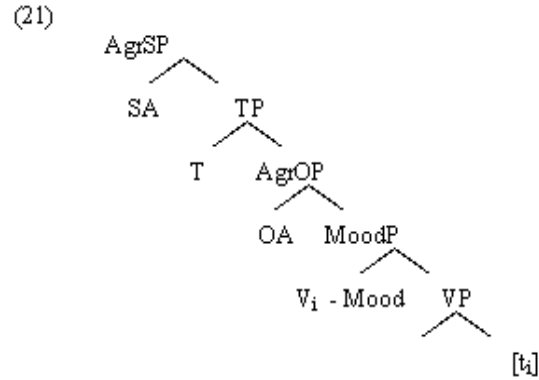
Let us summarize the predictions of the MOM:

- (20)
- I. SA should never be omitted (because it carries secondary stress).
 - II. T should never be omitted (because it is the weak syllable in a trochaic foot).
 - III. OA should always be subject to omission (because it is always a pre-trochaic weak syllable).

4.3.2 Truncation Hypothesis

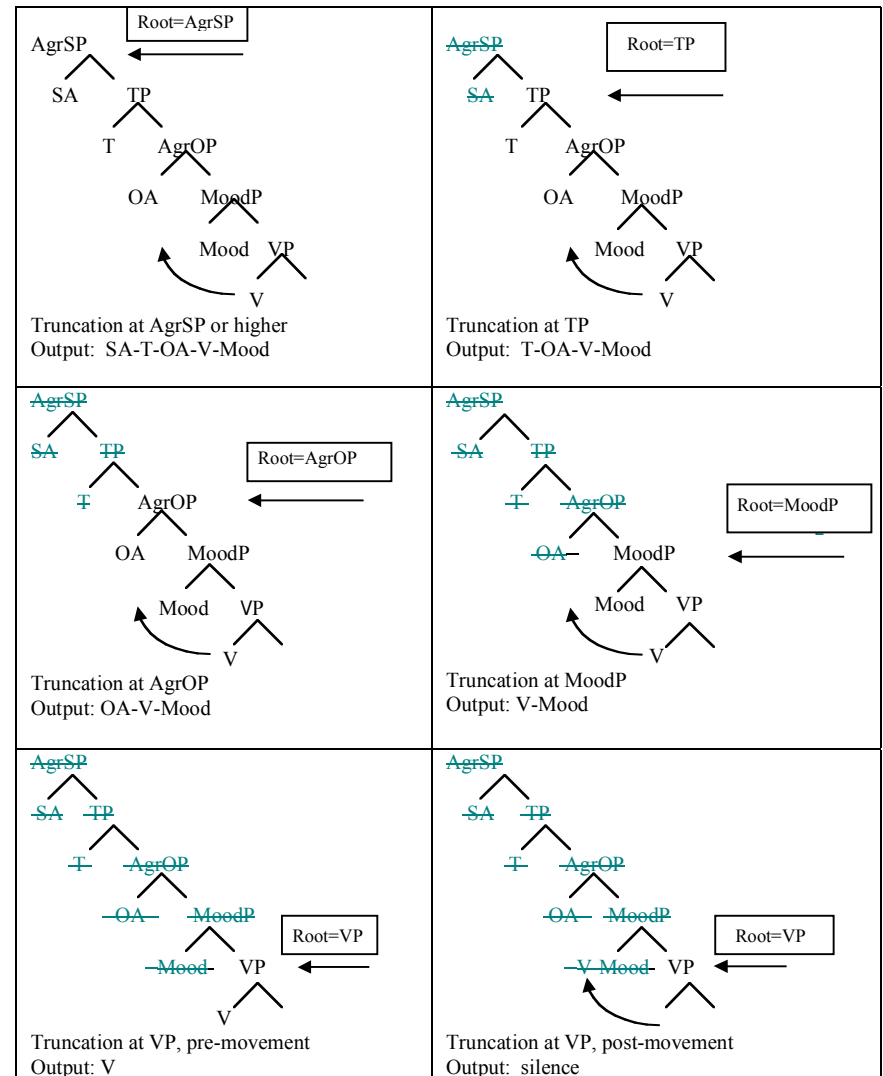
Swahili provides us with a language that is particularly clear in its morphological structure. Not only does it spell out three inflectional morphological heads as separate, independent morphemes, it does so in the order that corresponds to their hypothesized hierarchical order (Baker's, 1987 Mirror Theory; see Cinque, 1999 for discussion). Furthermore, there is very little phonological interference among inflectional morphemes, rendering surface forms fairly reliable indicators of underlying syntactic structure. Given this, Truncation provides a theory that has clear predictions with respect to early Swahili. Recall that truncation provides the child with the opportunity to specify the root as any projection in the syntactic tree. Let us assume the following structure of the Swahili verbal complex⁸¹:

⁸¹ In chapter 2 I gave a derivation of a Swahili simple tensed clause that involved head movement of the verb to Mood and remnant movement that re-ordered constituents into their surface order. I use a simpler structure here for the sake of exposition. Were we to use the more complex remnant structure the results would be the same. I also ignore specifier positions, but this again is for the sake of exposition.



In child grammar, according to Rizzi, the root may be specified as any projection above (and including) VP. Below are all the structures that the child could produce if truncation were to occur at each possible level. Truncation is silent on what level of representation truncation occurs at, i.e., it is not clear whether truncation occurs before movement or after movement. Therefore the status of Mood is unclear because prior to movement Mood is higher than the verb, but after the verb moves to Mood this is no longer the case. The only times this makes a difference is if truncation occurs at the VP level, in which case pre-movement truncation will yield a different result from post-movement truncation. I have given both possibilities below. I have indicated the “deleted” portion of the trees in lighter shade and strikethrough. Below each structure I give the output from that structure.

Recall that OA in Swahili is dependent on specificity of the object. This means that in all cases of a non-specific object, OA will not occur. Because of the difficulty in determining obligatory contexts for OA and for the sake of exposition, I will not consider it further. See section 4.3 for more details on OA in early Swahili.



Summarizing the results, we find that truncation predicts the following structures should be possible in early Swahili (ignoring the trivial option of silence).

- (22) Possible clauses according to truncation:
- | | | |
|------|------------|--------------|
| I. | SA-T-V-IND | Full Clause |
| II. | Ø-T-V-IND | [-SA] clause |
| III. | Ø-Ø-V-IND | Bare Stem |
| V. | Ø-Ø-V-Ø | Bare Root |

Furthermore, Truncation makes predictions regarding what sorts of clauses should NEVER occur in Swahili. Recall that while the root may be specified as any projection in the structure, all projections below the root *must* be projected – the child cannot omit any intervening material.

Therefore the following clause types are excluded by truncation:

- (23) Excluded clause types, according to truncation:
- | | | |
|------|------------|---------------------------|
| I. | SA-Ø-V-IND | [-T] clause with SA |
| II. | SA-T-V-Ø | [-Mood] clause |
| III. | SA-Ø-V-Ø | [-T] clause without Mood |
| IV. | Ø-T-V-Ø | [-SA] clause without Mood |

[-T] clauses are excluded as well as any clause that is missing mood but has a prefix. Thus mood may only be omitted when all other prefixes are omitted as well. Before moving on, we should note that supporting evidence for Truncation would include some or all of the possible clauses in (22) being attested, i.e., not all of the clause types in (22) need be attested. However, crucially, NONE of the excluded clauses in (23) can be attested. Any substantial (non-speech error) occurrence of the excluded clauses in (23) constitutes counter-evidence for Truncation.

4.3.3 Underspecification Theories

The underspecification theories make simple predictions. Wexler's underspecification of Tense theory predicts that children should allow the following clause types:

- (24) Permissible Clause types, according to Wexler (1994):
- | | |
|-----|------------|
| I. | SA-T-V-IND |
| II. | SA-Ø-V-IND |

Crucially, Wexler's theory does not allow SA to be omitted, and it does not allow Mood to be omitted. On the other hand, Clahsen *et al.*'s underspecification of Agr theory predicts a very different result:

- (25) Permissible Clause types, according to Clahsen *et al.* (1996):
- | | |
|-----|------------|
| I. | SA-T-V-IND |
| II. | Ø-T-V-IND |

Again, this underspecification theory predicts that SA may be omitted optionally, but T may not be omitted and Mood may not be omitted either. Finally, Schütze's (1997) ATOM predicts that the following options should be possible:

- (26) Permissible Clause types, according to Schütze's (1997) ATOM:
- | | |
|------|------------|
| I. | SA-T-V-IND |
| II. | Ø-T-V-IND |
| III. | SA-Ø-V-IND |
| IV. | Ø-Ø-V-IND |

ATOM predicts that SA and T may be optionally and independently omitted, but that Mood may not be omitted.

Table 4.2 Summary of Predictions

MOM:	SA and T should never be omitted, but OA and onset syllables of long verb stems are subject to omission.
Truncation:	Tense should never be omitted when SA occurs, and Mood should never be omitted when any prefixes occur. Other clause types are permitted.
Underspecification of T:	The only clauses permitted are full clauses and [-T] clauses.
Underspecification of Agr:	The only clauses permitted are full clauses and [-SA] clauses
ATOM:	Mood omission is excluded, otherwise all four combinations of prefix omission are permitted: Full clause, [-SA] clause, [-T] clause, Bare Stem

On more general grounds, based on what we know about children cross-linguistically, we might expect Swahili children to do two things:

- omit affixes
- Produce RIs

What would these look like in child Swahili? Let us consider the RI possibility first. Swahili has an infinitive marker that occurs in the position of T: *ku*. Recall from chapter 2 that the infinitive occurs in complementary distribution with all other tense markers, occurs in typical infinitival contexts (embedded clauses, gerunds, etc.), licenses PRO, etc. A child RI would include the infinitive marker, an indicative final mood vowel, but no SA (as SA and *ku* do not cooccur in adult Swahili). Thus we may schematize an RI as follows:

(27) INF-V-IND

Turning to possible omission of affixes, below are some possible clause types that we expect Swahili children to produce as well as labels that I assign the particular clause types. Because OA and the grammatical

function changing suffixes are optional, I ignore them here. I also do not consider mood omission because mood is less relevant to the theories that we discussed earlier.⁸² We will return to mood omission in section 4.6. Furthermore, I consider only indicative clauses in this chapter because subjunctive clauses have a different structure: tense is obligatorily absent in subjunctive clauses. The relative omissions of SA and T are crucial to our evaluation of the theories presented earlier (especially the underspecification theories). Because subjunctive clauses do not contribute to this discussion I will not consider them in this chapter. See Deen & Hyams (2002) for details on the use of subjunctive and indicative clauses.

(28) Logically Possible Omission types in child Swahili:

SA – T – V – Mood	Full Clause
∅ – T – V – Mood	[-SA] Clause
SA – ∅ – V – Mood	[-T] Clause
∅ – ∅ – V – Mood	Bare Stem

Thus these four clause types plus RIs are all logical possibilities in early Swahili. Of these four clause types, full clauses are the only clause type that is permitted in prescriptive Standard Swahili, while Nairobi Swahili permits full clauses and [-SA] clauses. Recall from chapter 2 that [-T] clauses and bare stems are virtually unattested in adult speech in indicative contexts (recall that imperatives occur as bare verbs. For this reason I will exclude imperatives in my analyses; see below).

⁸² Mood omission could in principle occur with each of the clause types given in the text. Therefore we can also consider [-Mood] full clauses, [-SA, -Mood] Clauses, [-T, -Mood] clauses, [-Mood] Bare stems, etc. We will see very soon that Mood is in fact almost never omitted. Thus, because these possibilities are unattested, for the sake of avoiding additional complexity I do not consider them.

In the next section I will present the results of analyses of omission of prefixes. In section 4.5 we will return to these theories and evaluate their predictions with respect to the Swahili data.

4.4 Results

In this section I will present the results in the following order. First we will see which clause types are attested in the corpus. I will discuss the details of categorizing utterances into these discrete classes, and how conflicting criteria were resolved. I will then discuss in detail the development of each clause type across stages. We will see that, quite expectedly, unadult-like clause types diminish across time, while the adult-like clause types increase in proportion. However, the relative proportions of each clause type and the differential rates of development are of great interest. In section 4.5 we will move to the evaluation of the five theories of language acquisition presented in section 4.2. Finally, we conclude with section 4.6.

4.4.1 Attested Clause Types in Child Swahili - Overview

CLAN programs were used to run automated counts of the clause types. The program COMBO was used because it allows for searches that combine text, e.g. it allows for a search of the combination of the codes SA and T, or SA and Ø, etc. A first level of analysis revealed the following facts:

- (29) General facts about early Swahili:
- SA omission is extremely frequent
 - T omission is also frequent
 - Mood Omission is almost completely absent
 - RIs are almost completely absent

In the entire corpus, the mood final vowel is omitted a total of 12 times out of a total of 2662 verbal utterances (an omission rate of 0.45%). Upon inspection, these 12 tokens are clearly speech errors because they usually occur in broken speech (i.e., when communication breaks down and the child is struggling for her next utterance), or the final vowel is assimilated to the following vowel making it difficult to determine whether it is present or not. The twelve examples are listed in appendix 4A. Thus I conclude that the omission of mood is unattested, and I will not consider it further.

Similarly, RIs are virtually unattested. Of all the 2662 utterances, there are a total of 14 RIs, which represents an RI rate of 0.52%. These 14 RIs are presented in appendix 4B. I assume that RIs are not part of the grammar of early Swahili, a fact that represents a marked difference between Swahili and languages such as German, French, Dutch, etc.

Having eliminated mood from consideration (although see Deen & Hyams, 2002), and concluded that RIs are not permissible in early Swahili, the only remaining possibilities are listed in (28) above. Again, automated CLAN programs were used to count the number of each clause type, and the results are presented below. I first present the data by child in separate tables, after which I present the data by stage. For a month-by-month breakdown of the clause types for each child, see appendix 4C.

Table 4.3 Clause types in Hawa's files

Clause Type	Frequency
Full Clause	13 (13%)
[-SA] Clause	20 (20%)
[-T] Clause	18 (18%)
Bare Stem	47 (48%)
Total	98

Table 4.4 Clause types in Mustafa's files

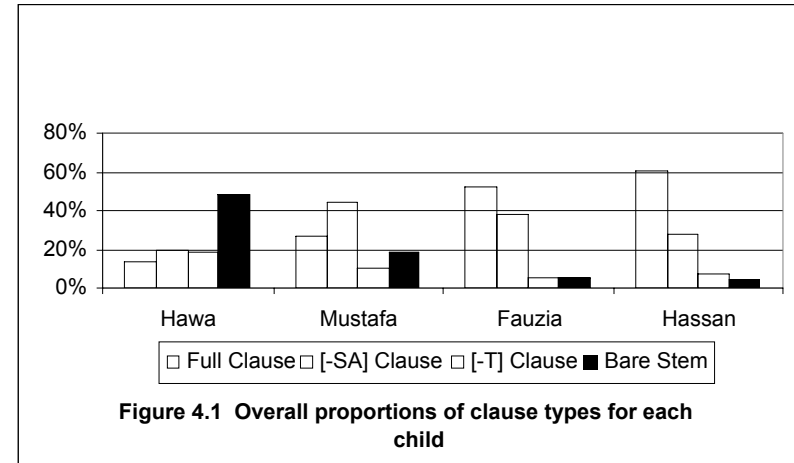
Clause Type	Frequency
Full Clause	136 (27%)
[-SA] Clause	225 (44%)
[-T] Clause	53 (10%)
Bare Stem	92 (18%)
Total	506

Table 4.5 Clause types in Fauzia's files

Clause Type	Frequency
Full Clause	183 (52%)
[-SA] Clause	135 (38%)
[-T] Clause	17 (5%)
Bare Stem	17 (5%)
Total	352

Table 4.6 Clause types in Hassan's files

Clause Type	Frequency
Full Clause	225 (61%)
[-SA] Clause	104 (28%)
[-T] Clause	26 (7%)
Bare Stem	15 (4%)
Total	370



As tables 4.3 – 4.6 show, all four possible clause types are attested in early Swahili (for examples of each clause type, see the sections on each clause type below). Let us start with a comparison of the four children. Recall we saw that Hawa is the least mature of the four children based on our three measures of linguistic maturity (MLU, V Ratio, %MPH), followed by Mustafa, Fauzia and Hassan. Beginning with Hawa in table 4.3, we see that the most frequent clause type in her files is the bare stem. Surprisingly, full clauses are the least frequent. This is surprising because full clauses are the most common indicative clause type in adult Swahili. We saw in chapter 2 that 95% of all indicative clauses in Swahili are full clauses. The other two clause types are represented in Hawa's data at between 18% and 20%, which is not insignificant. Turning to Mustafa, we see that bare stems are not the most frequent clause type in his files. Rather, Mustafa uses more [-SA] clauses than any other clause type. Furthermore, the second most common clause type is the full clause. These two clause types are the

only two that are permitted in the adult language. If we assume that Mustafa is at a stage in acquisition which is in advance of where Hawa is, then Mustafa is exhibiting a movement towards the adult target. These two clause types together represent 71% of all his indicative clauses (as compared to 33% for Hawa).

The remaining two children continue this move towards the adult norm in that they produce more full clauses than any other clause type. The second most frequent clause type in their files is the [-SA] clause, with full clauses and [-SA] clauses combining for 89% and 90% for Hassan and Fauzia respectively. However, while adults permit [-SA] clauses only 5% of the time, they occur in the speech of these children at rates of 28% and 38%. The trend then is that the less mature children are less adult-like in the clause types that they permit, while the more mature children are not fully adult-like, but closer to the adult norm than the others. This validates our staging procedure, because the children that we placed in lower stages based on three different measures exhibit less mature grammars than those placed in higher stages.

Let us turn to an analysis of the data by stage. This will unify the figures from tables 4.3-4.6 and allow us to see the development of each clause type in a much clearer fashion. Below are the frequencies of the four clause types by stage in table 4.7, followed by the proportions by stage⁸³:

⁸³ Monosyllabic placeholder used in unambiguous position (e.g., MPH-T-V-IND) were excluded from these counts. I determined whether the prefix was an MPH or a fully formed adult-like prefix based on context and pronunciation (see chapter 3). If an utterance included an MPH, it was not included in any of these counts.

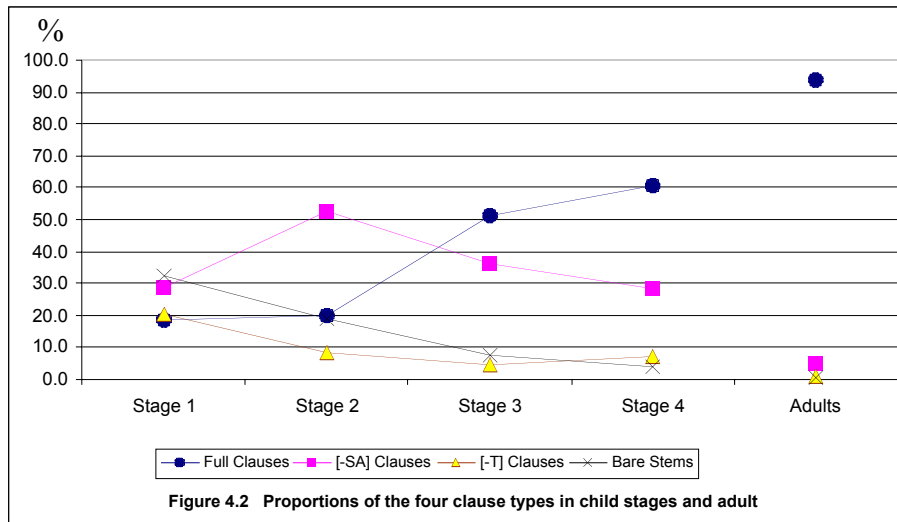
Table 4.7 Frequencies of clause types, by stage

	Stage 1	Stage 2	Stage 3	Stage 4	Total
Files from	Hawa 2;2-2;6, Mus 2;0-2;3	Mus 2;4-2;8	Fau 1;8-2;2, Mus 2;9-2;10	Has 2;10-2;11	
Full Clauses	39	58	235	225	557
[-SA] Clauses	60	154	166	104	484
[-T] Clauses	42	25	21	26	114
Bare Stems	67	55	34	15	171
Total	208	292	456	370	1326

Table 4.8 Proportions of clause types, by stage

	Stage 1	Stage 2	Stage 3	Stage 4
Files from	Hawa 2;2-2;6, Mus 2;0-2;3	Mus 2;4-2;8	Fau 1;8-2;2, Mus 2;9-2;10	Has 2;10-2;11
Full Clauses	18.8%	19.9%	51.5%	60.8%
[-SA] Clauses	28.8%	52.7%	36.4%	28.1%
[-T] Clauses	20.2%	8.6%	4.6%	7.0%
Bare Stems	32.2%	18.8%	7.5%	4.1%
Total	208	292	456	370

These proportions in table 4.8 are the sum of each clause type in a particular stage divided by the total number of indicative clauses in that stage (i.e., they pool data across children). We see that the description given earlier about the less mature children permitting more unadult-like clauses is starkly evident in the staged data in table 4.7. Below is a line graph of the proportions. For comparison, I have included the relevant proportions from the adults' speech in this corpus (proportions taken from chapter 2) on the right-hand side of the figure.



4.4.2 A note about Object Agreement

Before we consider each clause type in detail, a note about OA is required at this point. As discussed in chapter 2, OA is dependent on the specificity of the object: when the object is specific, OA is obligatory, but when the object is non-specific, OA is obligatorily absent. This raises a problem for determining obligatory contexts for OA because it is difficult to determine from context whether a child intends a specific or non-specific reading. However, some tentative findings are presented below. We will present results from two analyses that both suggest that children have knowledge of OA. In fact, the results will suggest that children's knowledge of OA may be ahead of their knowledge of SA inasmuch as we can tell from naturalistic data.

Each child did use OA, as table 4.9 shows:

Table 4.9 Number of OA in each child's files

Child	Number of OA
Hawa	35
Mustafa	40
Fauzia	95
Hassan	65
Total	235

The overall number of OA is lower than SA in the child files. A raw frequency count of the number of SA tokens in the child files reveals that SA occurs 978 times, while OA occurs 235 times. These numbers are not very meaningful because OA can only occur in transitive clauses, whereas SA can occur in all clauses. Children produced a total of 957 intransitive clauses and 1605 transitive clauses, as shown in tables 4.10-4.11:

Table 4.10 Total number of transitive clauses

Child	Number of transitive clauses
Hawa	283
Mustafa	557
Fauzia	466
Hassan	299
Total	1605

Table 4.11 Total number of intransitive clauses

Child	Number of intransitive clauses
Hawa	184
Mustafa	454
Fauzia	151
Hassan	168
Total	957

I first investigated whether children know that OA is only possible in transitive clauses⁸⁴. I calculated the proportion of OA that occurred in transitive clauses and the proportion of OA that occurred in intransitive clauses, and the results are presented below⁸⁵.

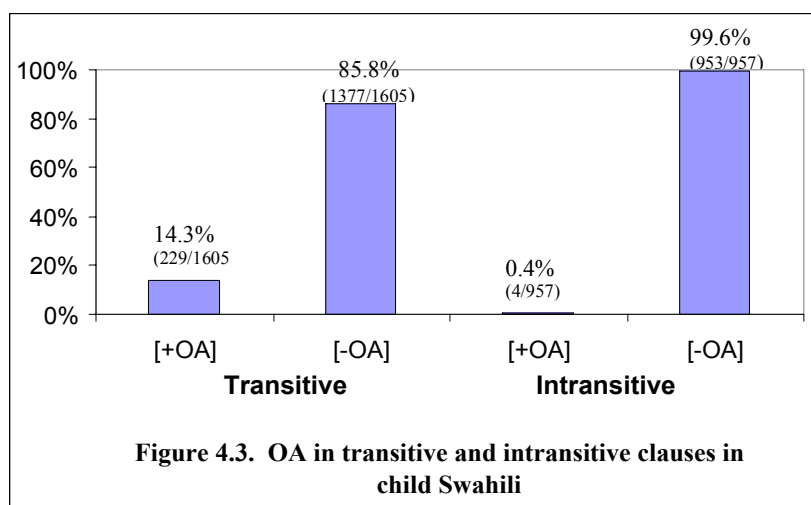
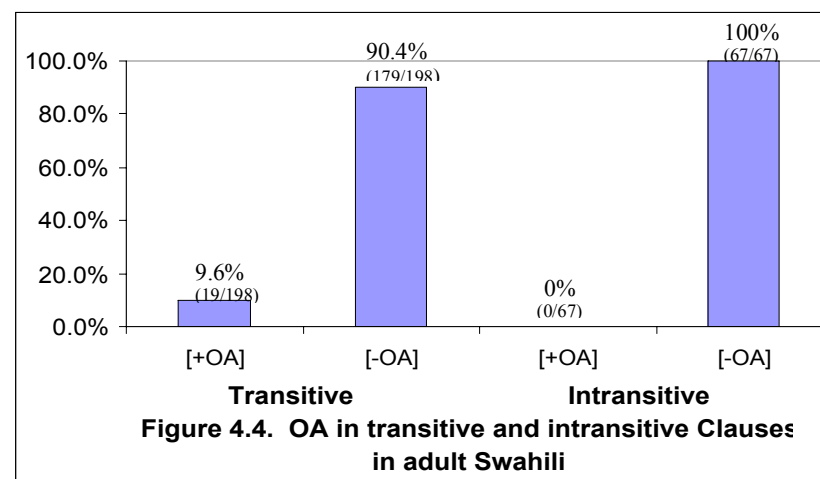


Figure 4.3 shows that while OA may be relatively infrequent in child Swahili, when OA occurs it is never on intransitive verbs. From this we can conclude that they at least know the function of OA in Swahili, that it is concerned with the presence of an object. As comparison, below is the figure for the adults in this corpus. Notice the striking similarity in

⁸⁴I used the subcategorization frame of the verb to determine transitivity, since object drop of overt object DPs is grammatical in Swahili.

⁸⁵The number of OA presented in this figure do not add up to 235 because this does not include distransitive clauses. When distransitive clauses are added in, the proportions are not greatly affected.

proportions. Not only are intransitives unmarked for OA, but OA in transitive clauses occurs at a similar rate as in child language.



The second analysis involves OA in contexts in which OA is obligatory. As mentioned earlier, obligatory context for OA is extremely difficult to determine from context, and so a comprehensive analysis of OA is not possible. However, the object may be obligatorily specific in some cases, e.g., when the object is a proper name.

The coding system unfortunately did not distinguish the grammatical roles of names, i.e., proper names were coded as names, with no distinction made between names that occur as subjects, names that occur as objects, names that occur as adjuncts, or names that occur as vocatives.

The total number of names in the corpus is 963, of which 189 are

postverbal⁸⁶. These 189 utterances were inspected by hand, and it was found that only 27 of these 189 are unambiguously objects that require OA (the majority of the remaining names were vocative, easily identified by intonation). Of these 27, a total of 25 carry the appropriate OA (92.6%). The remaining two clauses that are missing OA are reproduced below:

(29)

- a. h – u – Ø – beb – a Fauzia ? Has02, line 1012
 target: h – u – m – beb – a Fauzia?
 neg-SA_{2s}-OA_{3s}-carry-IND Fauzia
 ‘You don’t carry Fauzia?’
- b. Ø – on – a huyu Madua Fau07, line 1232
 target: m – on – a huyu Madua
 OA_{3s}-see-IND him Madua
 ‘See (him) Madua.’

In examples (29a-b) the OA has been omitted in obligatory contexts.

Example (29a) is a negative utterance in which the object is the name Fauzia, and (29b) is an imperative.

The overall numbers are not very large, and so more data is required in order to verify this result. However, if these results are shown to be valid, we must conclude the following. First, Swahili children never use OA in contexts in which it should not be used (i.e., intransitive clauses). Thus errors of commission involving OA are unattested. Second, errors of omission involving OA are also rare. There are only two cases of OA omission out of 27 obligatory contexts – an omission rate of 7.4%. I acknowledge that the frequencies are small and so I reserve full judgment

⁸⁶ Limiting the inspection to postverbal context was done for the sake of convenience. Children at this age do topicalize objects, and so this undoubtedly resulted in a reduction in tokens counted.

on this issue, but I present the data simply as suggestive and as a point of comparison with SA omission.

4.4.3 Full Clauses

Full clauses (FCs) are easy to identify in the corpus: all verbs that have SA and T. This includes those with and without OA. FCs occur with a full range of tense markers (examples 30a-f) as well as a wide range of SA (1,2,3 singular, and 1,3 plural), as well as more complex constructions. Example (30h) is a complex tense construction (grammatical in adult Swahili) and is composed of an auxiliary verb *kuwa* ‘to be’ that is fully inflected with SA and T, followed by a main verb that is also fully inflected with SA and T⁸⁷. Example (30i) is a biclausal utterance with two fully inflected clauses (grammatical in adult Swahili).

(30)

- a. ni – na – zim – a tuu ile Haw06, line 1160
 SA_{1s}-pres-finish-IND only that
 ‘I just finish that’ (Present tense)
- b. a – ta – la – a Mus10, line 1512
 SA_{3s}-fut-sleep-IND
 ‘She will sleep’ (Future tense)
- c. mimi ...ni – li – tow – a Mus12, line 1546
 I SA_{1s}-past-remove-IND
 ‘I...removed (it)’ (Past tense)

⁸⁷ Notice that the SA on both verbs are identical, an interesting feature of complex tenses that argues for an agreement analysis of SA since it would be unusual to have two identical subjects in the same clause. See chapter 2 for more discussion on this point.

- d. sku gani u – ta – end – a ? Fau03, line 226
 day which SA_{2s}–fut–go–IND (Future tense)
 ‘On what day will you go?’
- e. ah weh u – sha – um – a Fau07, line 1090
 you SA_{2s}–pr.prf.–hurt–IND (Past perfective)
 ‘Ah, you have already hurt (me)’
- f. a – ka – end – a Has01, line 547
 SA_{3s}–cont–go–IND (Continuative tense)
 ‘He (then) went.’
- g. wa – na – kul – a pilau Has01, line 485
 SA_{3pl}–pres–eat–IND pilau (Plural SA)
 ‘They eat Pilau.’
- h. tu – li – kuw – a tu – na – kul – a pilau! Has01, line 794
 SA_{1pl}–past–be–IND SA_{1pl}–pres–eat–IND pilau (Plural SA and
 ‘We were eating Pilau.’ complex tense)
 (lit: We were we are eating Pilau)
- i. Baba, u – li – sem – a Has01, line 333
 Father SA_{2s}–past–say–IND (Embedded clause)
- u – ta – nunu – a nini ?
 SA_{2s}–fut–buy–IND what
- ‘Father, what did you say you will buy?’
 (lit: Father, you said you will buy what?)

Each of these examples is completely adult-like. Occasionally utterances occur that have a reduced SA, as in example (31) below, where [ni] is reduced to a syllabic nasal. Such cases were categorized as FCs because in each such case the reduction was consistent with adult spoken language.

- (31) n – na – on – a Haw06, line 1720
 prefix – pres – see – IND
 unreduced target: ni – na – on – a
 SA_{1s} – pres – see – IND
 ‘I see.’

Furthermore, no cases occurred in which the prefix did not phonologically resemble the well-formed target. For example, no cases of [ɲ] or [∅] were used instead of [ni]. Therefore such cases of reduction were not counted as MPHs, but as full SA prefixes.

Finally, FCs occur with overt as well as null subjects, with and without OA, and they occur with grammatical function changing suffixes. In this respect, they are adult-like (see sections 2.4.4 (OA), 2.4.6 (suffixes) and 2.9 (null subjects) for discussion of adult Swahili).

4.4.4 [-SA] Clauses

[-SA] clauses are the most frequent clause types in the early stages, accounting for more than half of all the indicative clauses in stage 2. There are several features of [-SA] clauses that I will illustrate: They occur with overt subjects as well as null subjects; with OA and without OA; and with grammatical function changing suffixes such as applicative or passive. In each case [-SA] clauses occur with various T markers.

[-SA] clauses occur with overt subjects (32) as well as with null subjects (33). For a full discussion of this phenomenon, see chapter 5. Irrespective of whether the subject is overt or null, a wide range of tense markers occur in these clauses.

(32) **Overt subjects:**

a. mimi Ø – na – ruk – a Mus22, line 478
target: mimi ni – na – ruk – a (Present tense)
I SA_{1s}-pres-jump-IND
'I am jump down.'

b. baba Ø – li – tok – a ku – jeng – a Has02, line 255
target: Baba a – li – tok – a ku – jeng – a (Past tense)
Baba SA_{3s}-past-leave-IND inf-build-IND
'Father left to build'
(Father is a construction worker)

c. mimi Ø – li – m – chap – a Fau04, line 116
target: mimi ni – li – m – chap – a (Past tense)
I SA_{1s}-past-OA_{3s}-slap-IND
'I slapped him.'

d. ndege Ø – me – end – a juu Mus21, line 172
target: ndege i – me – end – a juu (Pres.perfect)
bird SA_{inanim}-p.perf.-go-IND up
'The bird has gone up.'

e. mimi Ø – ta – ingi – a apa ndani Fau08, line 160
target: mimi ni – ta – ingi – a hapa ndani (future tense)
I SA_{1s}-fut-enter-IND here inside
'I will enter here inside.'

(33) **Null Subject:**

a. Ø – na – lal – a Haw07, line 681
target: ni – na – lala – a (Present tense)
SA_{1s}-pres-sleep-IND
'I am sleeping.'

b. Ø – li – tumbu – a Mus18, line 378
target: a – li – ni – sumbu – a (Past tense)
SA_{3s}-past-OA_{1s}-disturb-IND
'He disturbed me.'

d. Ø – ta – po – a Mus20, line 431
target: ni – ta – po – a (Future tense)
SA_{1s}-fut-relax-IND

'I will relax.'

e. Ø – me – tok – a juu Mus21, line 155
target: i – me – tok – a juu (Pres. Perfect)
SA_{inanim}-p.perf.-leave-IND up

'It has left (and gone up).'

f. Ø – ka – nunu – a soda Fau09, line 1015
target: a – ka – nunu – a soda (Continuative)
SA_{3s}-cont-buy-IND soda

'He then bought soda.'

OA occurs in [-SA] clauses, although as mentioned earlier, OA in general is less common than SA. Therefore the variety of tense markers that we see is not as wide.

(34)

a. Ø – ta – ku – pig – a Mus20, line 135
target: ni – ta – ku – pig – a
SA_{1s}-fut-OA_{2s}-hit-IND
'I will hit you.'

b. Ø – ta – i – pig – a Fau08, line 801
target: ni – ta – i – pig – a
SA_{1s}-fut-OA_{inanim}-hit-IND
'I will hit it.'

c. Ø – ta – ni – p – i – a mimi? Has03, line 1624
target: u – ta – ni – p – i – a mimi?
SA_{2s}-fut-OA_{1s}-give-applic-IND me
'Will you give me (it)?'

Finally, [-SA] clauses occur with grammatical function changing suffixes such as the applicative (35a and 35b) and the passive (35c and 35d).

Example (35e) is the only case of a [-SA] clause occurring with stacked

suffixes. This extremely complex example has an applicative, causative and a passive suffix.

- (35)
- a. Ø – ta – nunu – li – a hii soda? Fau06, line 528
 target: u – ta – nunu – li – a hii soda?
 SA_{2s}–fut–buy–applic–IND this soda
 ‘Will you buy (for me) this soda?’
- b. weh Ø – na – imb – i – a ...kalamba Fau09, line 1096
 target: Wewe u – na – imb – i – a Kalamba?
 You SA_{2s}–pres–sing–applic–IND Kalamba
 ‘Will you sing Kalamba (for me)?’
- c. Ø – na – it – w – a Tafa Fau03, line 808
 target: a – na – it – w – a Tafa
 SA_{3s}–pres – call – passive – IND Tafa
 ‘He is called Tafa.’
- d. gubi Ø – me – dung – w – a Mbize? Mus23, line 46
 target: Kumbikumbi i – me – dung – w – a Mbize?
 White ant SA_{3s}–p.perf.–dig–passive–IND Mbize
 ‘Has the white ant dug (into the ground), Mbize?’
- e. Ø – na – vaa – li – sh – w – a Fau08, line 1649
 target: a – na – vaa – li – sh – w – a
 SA_{3s}–pres–wear–applic–causative–passive–IND
 ‘He is dressed’
 (lit.: He is caused to have the clothes to be put on.)

The majority of [-SA] clauses, however, are simple mono-prefixal clauses (i.e., tense) with no suffixes other than the mood final vowel. In some cases there was some difficulty in categorizing an utterance as either a [-SA] clause or something else. Recall from section chapter 3 that Swahili children produce MPHs (mono-syllabic place holders, Bottari *et al.*, 1993).

Refer to section chapter 3 for a discussion on how MPHs were distinguished from other well-formed prefixes.

Turning now to the development of [-SA] clauses, we see that in stage 1 they occur at 28.8%. They increase in stage 2 to over 50% and then diminish, but by stage 4 they are still very common, with 28.1% of all indicative clauses missing SA. Notice that the starting proportion (28.8%) and the final proportion of [-SA] clauses (28.1%) is approximately the same. There are several way to interpret this data, but one thing that this suggests is that while there is a surge in the number of [-SA] clauses in stage 2, this may be attributable to factors other than development. For example, it may be that Mustafa (the only child in stage 2) uses an unusually high proportion of [-SA] clauses. Indeed, a count of Mustafa’s files reveals that [-SA] clauses (across all 11 months) account for 44.5% (225/506) of all his indicative clauses. This compared to 18% for Hawa, 38% for Fauzia and 28% for Hassan. Therefore, while the staging process is useful in allowing us to see overall trends, we must take into account individual variation among children. The conclusion is that [-SA] clauses show little development across the four stages. In adult Swahili [-SA] clauses occur at a rate of approximately 5%, which is in stark contrast to child Swahili (at any stage or for any child). Therefore the principles that are responsible for the reduction of [-SA] clauses in child language develop at a stage beyond the last files in this corpus. This is an area for future research.

4.4.5 [-T] Clauses

[-T] clauses are the least common clause type amongst the four, and are the first to diminish to under 10%. They begin at a rate of over 20%,

but immediately fall to under 10% in stage 2, and remain under 10% through to stage 4. Therefore I consider stage 2 the point at which [-T] clauses cease to be a possibility in early Swahili. Recall that [-T] clauses are completely ungrammatical in adult Swahili, not occurring in naturalistic discourse nor being accepted in elicitation by native speakers (see chapter 2).

Although a few [-T] clauses occur with OA (as in example (36) below), they primarily occur without OA.

(36) a – Ø – mw – on – a Haw06, line 1613
 target form: a – na – mw – on – a
 SA_{3s}–pres– OA_{3s}–see–IND
 ‘He sees her.’

Categorizing utterances as [-T] clauses occurred through a process of elimination. [-T] clauses occur with a single prefix, and so there is a possibility of misclassifying a [-T] clause as [-SA], a clause with an MPH, or a subjunctive clause with SA as the single prefix:

(37) a. [-T] clause: SA – V – IND
 b. [-SA] clause: T – V – IND
 c. MPH clause: prefix – V – IND
 d. Subjunctive: SA – V – SUBJ

For the most part, [-SA] clauses occurred with a single syllabic prefix on the verb. In most cases the verb stem itself was marked for mood and nothing else (i.e., no additional suffixes such as applicative, passive, etc.). Categorizing utterances as [-SA] or [-T] was relatively easy since the respective morphemes are very distinguishable. However, distinguishing between [-T] clauses and verbs with an MPH prefix is sometimes more difficult because there are two SA markers that are vowels with no onset consonant (2nd person singular [u] and 3rd person singular [a]). However, MPHs are distinguished from SA markers by the fact that they are reduced

vowels or single nasals, none of which occur as SA (see earlier discussion of MPHs and the method used to determine whether a prefix is an MPH or not).

[-T] clauses have the morphological structure given in (37a) above. They contain SA immediately followed by the verb root and the indicative mood final vowel ([a]). A morphologically similar clause is the subjunctive, schematized in (37d). Notice that the only difference between these two clause types is the final vowel ([a] for indicative, [e] for subjunctive). I have been considering clauses of the kind in (37a) as full clauses with a missing T. However, it is possible that they are in fact subjunctive clauses with a mispronounced final vowel.

If [-T] clauses are subjunctive clauses with a mispronounced final vowel, this represents an error of commission. Recall that errors of commission are very rare in child language in general. On the basis of what we know about other child languages, it is unlikely that children mispronounce the final vowel in [-T] clauses. So on conceptual grounds we reject this analysis, and I show empirically below that this cannot be the case.

Deen & Hyams (2002) show that children use indicative/subjunctive morphology correctly in Swahili. In all the indicative clauses in this corpus, children misuse indicative morphology 23/1436 (1.6%) times. In other words, 1.6% of verbs that are marked as indicative are used in subjunctive contexts. In adult Swahili, indicative marking co-occurs with temporal marking. Because of this relation between mood and tense, we speculate that this low percentage of mood errors in child Swahili may be due to the presence of T in some of those cases (i.e., T may be forcing a temporal, non-subjunctive interpretation), and so we calculate the

proportion of misused indicative morphology in verbs that are missing T: bare stems. Of 164 bare stems that are marked indicative, 5 occur in subjunctive contexts (3%). Thus it is not the case that the presence of T forces an indicative interpretation. Our conclusion is that Swahili children do not misuse indicative morphology when they intend to use subjunctive morphology, whether T is present or not.

Similarly, of the 114 [-T] clauses, there are 16 cases in which it is unclear what the intention of the child was. Therefore these 16 cases are not considered in this particular analysis (but are included in the general proportions presented thus far). Of the remaining 98 [-T] clauses, 8 are compatible with a subjunctive-type meaning, i.e., they could have been intended as a request, or a suggestion, or an expression of possibility or desire, etc⁸⁸. This represents an ‘error’ rate of 8/98 (8.2%). When we combine these numbers with other clauses that are missing T (i.e., bare stems), the overall ‘error’ rate is 13/262 (4.9%). Thus I conclude that errors of this sort are rare. Therefore, because the interpretation of [-T] clauses is consistent with indicative morphology, I assume that [-T] clauses are not final vowel errors, but represent errors of omission in which T has been omitted from a full clause.

A variety of [-T] examples are given in (38) below. Examples (38a-c) are from Fauzia. Notice that each utterance has a different SA – 1st, 2nd, and 3rd person singular. Examples (38d-e) from Hawa also have different SA markers, as do the examples from Hassan and Mustafa. This shows that SA is used productively in [-T] clauses and reference is not limited to any particular person. Furthermore, a wide range of intended

temporal meanings are represented: present tense (38a, d, g and h), past (38b, c, e, f and j), and present perfect (38i). Therefore temporal reference appears to be free in [-T] clauses.

- (38)
- | | | |
|---------|---|------------------|
| a. | ni – Ø – kw – ambi – a | Fau04, line 95 |
| target: | ni – na – kw – ambi – a
SA _{1s} –pres–OA _{2s} –tell–IND
‘I am telling you.’ | |
| b. | u – Ø – kingi – z – a ? | Fau07, line 1044 |
| target: | u – li – kimbi – z – a?
SA _{2s} –past–run–causative–IND
‘Did you make (him) run?’ | |
| c. | a – Ø – sem – a nini ? | Fau04, line 464 |
| target: | a – li – sem – a nini?
SA _{3s} –past–say–IND what
‘What did he say?’ | |
| d. | ni – Ø – chun – a | Haw05, line 682 |
| target: | ni – na – chun – a
SA _{1s} –past–pinch–IND
‘I am pinching!’ | |
| e. | alafu a – Ø – rud – i | Haw07, line 905 |
| target: | alafu a – li – rud – i
then SA _{3s} –past–return–IND
‘Then he returned.’ | |
| f. | alafu a – Ø – kuj – a | Has01, line 1178 |
| target: | alafu a – li – kuj – a
then SA _{3s} –past – come – IND
‘Then he came.’ | |
| g. | u – Ø – on – a nimó | Has01, line 1396 |
| target: | u – na – on – a Nimó?
SA _{2s} –pres – see – IND Nimó
‘Do you see Nimó?’ | |

⁸⁸ One had a desiderative interpretation, one was used as a request, two expressed deontic necessity, and four were suggestions.

h. u – Ø – gong – a hapa Mus04, line 1150
 target: u – na – gong – a hapa
 SA_{2s}– pres – knock – IND here
 ‘You are knocking here.’

i. ni – Ø – pit – a Mus05, line 316
 target: ni – me – pit – a
 SA_{1s}– p.perf. – pass – IND
 ‘I have passed (through).’

j. a – Ø – tup – a Mus03, line 2163
 target: a – li – tup – a
 SA_{3s}– past – throw – IND
 ‘He threw (something).’

As we will see in more detail, subjects generally do not occur in [-T] clauses. I will delay discussion of this point, and refer the reader to chapter 5 for a full discussion of the details and reasons for this.

4.4.6 Bare Stems

Bare stems are verbs that are missing SA and T, but have a mood final vowel. In stage 1 they are the most common clause type of the four attested indicative clause types. They quickly diminish, going from 32.3% in stage 1, to 18.8% in stage 2, to under 10% by stage 4. By stage 4 bare stems occur a mere 4% of the time. Examples of bare stem clauses from each child are presented below. In (39a-e) we have examples from Hawa’s files. Notice that the interpretation includes past, present and present

perfective. Also, the persons are 1st or 3rd person⁸⁹. The other children show similar characteristics, with example (39j) from Hassan’s files showing a continuative reading.

(39)

a. Ø – Ø – anguk – a hivi Haw07, line 210
 target: ni – li – anguk – a hivi
 SA_{1s}–past–fall–IND like this
 ‘I fell like this.’

b. Ø – Ø – tak – a tuwadh–a Haw07, line 5
 target: ni – na – tak – a tuwadh–a
 SA_{1s}–pres–want–IND bathe–IND
 ‘I want to bathe.’

c. Ø – Ø – ndik – a Haw05, line 2946
 target: ni – me – andik – a
 SA_{1s}–p.perf.–write–IND
 ‘I have written.’

d. Ø – Ø – ka – a hapa Haw04, line 130
 target: a – na – ka – a hapa
 SA_{3s}–pres–live–IND here
 ‘She lives here.’

e. Ø – Ø – end – a job Haw03, line 128
 target: a – me – end – a job
 SA_{3s}–p.perf.–go–IND job
 ‘He has gone to work.’

⁸⁹ Deen & Hyams (2002) find that the bare stems in Swahili do not have a modal meaning, parallel to English bare stems (Hoekstra & Hyams, 1998). Of the 164 bare stems in the corpus, only 5 (3%) occurred in contexts that were compatible with a modal meaning. This supports the idea that infinitival morphology licenses MoodP. Thus the irrealis meaning of RIs (Hoekstra & Hyams’ Modal Reference Effect) is not a result of a simple lack of temporal specification, but rather arises from the presence of infinitival morphology.

f.	aya, Ø – Ø – annguk – a	Mus19, line 930
target :	aya, ni – na – anguk – a SA _{1s} -pres – fall – IND 'Aya, I am falling.'	
g.	Ø – Ø – um – a	Mus17, line 632
target :	a – li – ni – um – a SA _{3s} -past - OA _{1s} -hurt – IND 'He hurt me.'	
h.	Ø – Ø – kata – a	Fau06, line 863
target:	ni – na – kata – a SA _{1s} -pres-refuse-IND 'I refuse.'	
i.	Ø – Ø – kasilik – a	Fau01, line 1709
target:	ni – me – kasirik – a SA _{1s} -p.perf.-get angry-IND 'I am angry.' (lit: I have gotten angry)	
j.	Ø – Ø – shuk – a chini	Has01, line 687
target:	a – ka – shuk – a chini SA _{3s} -cont-align-IND down 'He then got down.'	
k.	Ø – Ø – kul – a	Has03, line 855
target:	ni – na – kul – a SA _{1s} -pres-eat-IND 'I am eating.'	
l.	baba Ø – Ø – end – a hoa	Has03, line 1309
target:	Baba a – me – end – a hoa Baba SA _{3s} -p.perf.-go-IND hoa 'Father has gone home.'	

As we did for the other clause types, we will now consider the coding of bare stems. I have presented bare stems as having underspecified Tense and Agr. However, given the frequency of RIs in languages such as

German and French, it is possible that Swahili bare stems are actually RIs with the infinitival prefix missing. After all, an infinitival verb (INF-V-IND) would have the correct metrical structure so as to allow metrical omission of the INF prefix (assuming that SA is syntactically omitted).

The infinitive prefix, being in a preverbal position, would take weak stress. Being in a pre-trochaic position, it would thus be subject to omission by the MOM.

(40)	W S W	→ W [S – W]
syllabic:	ku-en-da	
morphemic:	ku – end – a	
	inf – go – IND	
	'to go'	

Wexler (1994) suggests that English bare verbs be assimilated to RIs. Under this view English bare stems and Swahili bare stems are considered RIs that appear as bare verbs because of language-specific morphology. The appeal of this idea is that it unifies the analysis of English bare verbs, Swahili bare stems and RIs. Under this view, children acquiring English produce bare verbs while German, Dutch and French children produce RIs because of morphological differences between these languages: the infinitive in English is not a true infinitive in that it is not a bound affix on the verb (material may occur between the verb and the particle: 'to boldly go where no one has gone before'). Furthermore, the infinitive marker in English is preverbal while in RI languages it is a suffix. Moreover, adult English uses bare verbs in contexts in which adult RI languages use infinitives, e.g., as complements to causatives, or perception verbs.⁹⁰

⁹⁰ Thanks to Carson Schütze for discussion on this point and the French example.

- (41) Il nous a fait chanter.
 French
 He us has made sing-infin
 'He made us sing.'

However, true RIs generally have a modal or irrealis meaning (Wijnen, 1987; Blom & Wijnen, 2001). This has been shown to be true for Dutch, French, German and Swedish, although the evidence for other RI languages is not so clear. However, English bare verbs, as noted by Hoekstra & Hyams (1998), tend to have a temporal meaning, usually present tense here-and-now, but also past tense (Deen, 1997; Madsen & Gilkerson, 1999; Torrence, 2002). Hoekstra & Hyams attribute the irrealis meaning associated with RIs to the presence of infinitival morphology, which they note, cross-linguistically is generally associated with irrealis or modal meaning in adult languages. Therefore there is a clear semantic difference between true RIs and English child bare verbs. This argues against an RI analysis of English bare verbs.

As we have seen in this section, Swahili children produce bare verbs, as English-acquiring children do. We saw in section 4.4.5 that the interpretation of bare verbs in Swahili is overwhelmingly non-modal (5/164 bare verbs occur in irrealis contexts), and thus Swahili bare verbs pattern with English bare verbs. Moreover, as in English, RIs are completely unattested in Swahili. Thus, at first glance Swahili appears to be parallel to English in many ways. However, unlike English, the infinitive in Swahili is a true infinitive. It is a bound affix on the verb. The only other material that can occur between it and the verb is OA, which is expected given that AgrOP occurs below TP. Finally, complements to perception verbs occur in the subjunctive or as fully inflected verbs, but not as bare verbs. Thus

Swahili does not have the same properties of English that purportedly result in bare verbs in child language.⁹¹

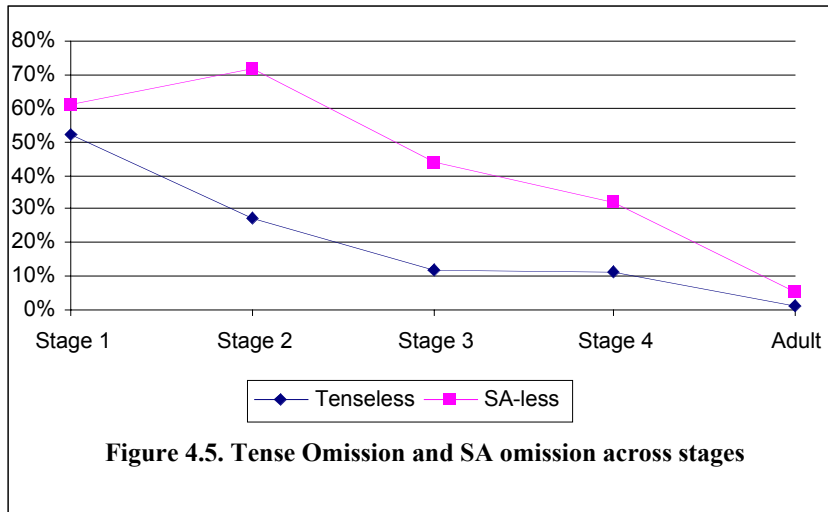
Hoekstra & Hyams (1998) showed that significant differences exist between RIs and bare verbs in child language. The Swahili facts presented here show that despite having a true infinitive in the adult language, bare verbs are a genuine possibility in child Swahili.

4.4.7 Tense Omission versus SA omission

Thus far we have seen that Swahili children permit four clause types: full clauses, [-SA] clauses, [-T] clauses and bare stems. We saw in chapter 2 that Swahili adults only allow two clause types: full clauses and [-SA] clauses. Therefore the crucial difference between adult Swahili and child Swahili is the omission of T. Let us therefore examine clauses that are missing T as a single class. Below is a graph showing the occurrence of tenseless clauses ([-T] clauses and bare stems) across the developmental

⁹¹ Additionally, several of the syntactic correlations that have been noted in RIs in true RI languages (German, Dutch, French, etc.) do not hold in English. For example, null subjects predominate in non-finite contexts in RI languages, but not in English (see section 1.2.3). Also, RIs generally do not occur in wh- questions in languages such as German and Dutch, but in English wh- questions occur with bare verbs (see Roeper & Rohrbacher 1994 and Bromberg & Wexler, 1995, who look at wh- fronting and null subjects in finite and non-finite contexts). Therefore English bare verbs differ from true RIs in some significant syntactic ways. These tests cannot be applied to Swahili because Swahili allows null subjects and wh- words remain in situ.

stages and the occurrence of those clauses in which SA is omitted (including [-SA] clauses and bare stems, labeled ‘SA-less’).⁹²



Overall, the omission of SA and the omission of T tend to diminish as the children mature. However, tense omission ceases to be a possibility in early Swahili much earlier than SA omission. By stage 3, tense omission is rare (12%), while SA omission at that same stage occurs at a rate of more than 40%. Similarly while T omission in stage 4 is rare, SA omission is still above 30%. Another way to view this is that SA omission is tolerated longer than T omission. We will return to this point in the section in which

⁹² Including bare stems in both categories is necessary because I wish to compare SA omission on the one hand and T omission on the other. Bare stems fall into both categories.

we review ATOM’s (Schütze & Wexler, 1996) applicability to the Swahili data.

4.4.8 Summary

We investigated the occurrence of various verbal clauses in the speech of four Swahili speaking children, and found that not only do they allow clause types that adults allow, they also allow two clause types that are unattested in adult speech. We found that full clauses are used relatively sparingly at early stages in development, but develop into the majority clause type by stage 4. We also found that [-SA] clauses are used very frequently, and remain a genuine possibility beyond stage 4. We saw that the use of tense in these clauses is fully productive, and the reference of the missing SA is free. Overt subjects and null subjects occur in both full clauses as well as [-SA] clauses.

In addition to full clauses and [-SA] clauses (both of which are attested in adult Swahili), children produce verbal clauses that are missing T (which we call [-T] clauses) and verbal clauses that are missing both T as well as SA (which we call bare stems). We saw that [-T] clauses cease to be a possibility fairly early in development, falling to under 10% by stage 2. Bare stems also diminish, but do not fall to under 10% until stage 4. We saw that [-T] clauses allow a range of temporal references and occur with a range of SA markers, while bare stems allow for a range of temporal meanings as well as persons. In both cases we found that they were generally incompatible with a subjunctive-type meaning, and so we concluded that they were indicative clauses with missing T and/or SA.

Finally, we noted that T omission is unattested in adult Swahili, but is a possibility in child Swahili. SA omission, on the other hand, is

possible in both adult Swahili as well as child Swahili. Seemingly aware of this difference, Swahili children tolerate the omission of SA even into stage 4, but T omission ceases to be a significant phenomenon in stage 4.

In the next section, we will re-evaluate the theories presented in section 4.2 in light of what we now know about child Swahili. We will see that none of the theories fully explain everything that we have noted, but Schütze & Wexler’s (1996) ATOM captures the general facts of Swahili. There are several facts, however, which remain unexplained even by ATOM.

4.5 Revisiting Acquisition Theories

In section 4.2 we discussed five influential theories of the acquisition of early morphosyntax. We discussed the Metrical Omission Model which proposes that children omit syllables that occur in particular metrical configurations. We also discussed Rizzi’s (1994) Truncation Hypothesis, which proposes that children can optionally specify the root of the clause as a lower projection than CP. We then discussed three underspecification theories: an underspecification of T theory (Wexler, 1994), an underspecification of Agr theory (Clahsen *et al.* 1996) and a theory of underspecification of both T and Agr (Schütze & Wexler, 1996). We will evaluate each of these five theories in the following sections.

4.5.1 Metrical Omission Model

At first glance, the MOM appears to have promise. The morphemes that are optionally omitted in Swahili are individual syllables, and so can be individually targeted for metrical omission. The inflectional prefixes in Swahili are preverbal, and we know from chapter 2 that Swahili

is a trochaic language. However, recall our predictions from section 4.2.6. We saw that in verbal complexes with disyllabic verb stems, MOM predicts no omission (42a)⁹³. In verbal complexes with trisyllabic verb stems, MOM predicts omission of only the onset syllable in the verb stem⁹⁴. Therefore no omission of prefixes is predicted.

- (42) a.
$$\begin{array}{ccccccc} S & - & W & - & S & - & W & \rightarrow & [S - W] & [S - W] \\ \vdots & & \vdots & & \vdots & & \vdots & & & \\ ni & - & ta & - & pik & - & a & & (chakula) \\ SA_{1s} & - & fut & - & cook & - & IND & & (food) \\ 'I & & will & & cook & & food' \end{array}$$
- b.
$$\begin{array}{ccccccc} S & - & W & - & W & - & S & - & W & \rightarrow & [S - W] & W & [S - W] \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & & & \\ ni & - & ta & - & anguk & - & a \\ SA_{1s} & - & fut & - & fall & - & IND \\ 'I & & will & & fall' \end{array}$$

⁹³ One objection could be that the addition of OA may affect the parsing of the string such that metrical omission becomes relevant. However, the crucial fact is that SA carries secondary stress, and hence is a strong syllable. Two side-by-side strong syllables are dispreferred, and so because SA is a strong syllable, T is a weak syllable. Therefore SA and T will always form a trochaic foot. The addition of OA will only add a weak syllable between the trochaic foot of the prefixes and the verb stem. Therefore MOM may have predictive power for the omission of OA, but it cannot explain the omission of SA or T. The addition of any other affixes (e.g., the applicative suffix, the passive suffix, etc.) will only affect the metrical structure of the latter portion of the string and will leave SA and T in a trochaic foot.

⁹⁴ In fact, there is sporadic evidence that this may be correct: verbs such as *anguka* occur frequently as [uka] – the onset [a] being omitted. This occurs when the verb is bare, as well as when the verb has some or all prefixes. Thus this is evidence that MOM actually does apply as Gerken (1991) proposes, but it does not explain the omission of the inflectional prefixes.

This is plainly contradicted by the data. Not only are prefixes omitted by children, at stage 1 over 80% of indicative verbs are missing either SA, T or both. The overwhelming frequency of prefix omission thus is not compatible with MOM. This is in contrast to Demuth's (1994) conclusion that the MOM (at least in part) accounts for the omission of noun class prefixes in Sesotho.

4.5.2 Truncation Hypothesis

In section 4.2.6 we discussed the predictions that Truncation makes for Swahili and we listed the clause types that we expect if Truncation applies in early Swahili. We saw earlier that Mood is never omitted. In terms of Truncation we interpret this as truncation always occurring above Mood (see below). We also saw earlier that it is difficult to determine precisely whether OA is omitted or not (although the little evidence that there is suggests that it isn't omitted). Putting aside OA for this reason, Truncation makes the following predictions:

(43) Predictions for Truncation:

- (i) Full clauses should optionally occur
- (ii) [-SA] clauses should optionally occur
- (iii) Bare stems should optionally occur
- (iv) [-T] clauses should be impossible.

Clearly predictions (i) – (iii) hold, which provides evidence for Truncation. However, prediction (iv) is false, as we saw that in early stages [-T] clauses occur at substantial levels (over 20% in stage 1). Truncation can not account for [-T] clauses because they contain SA (which necessarily entails that the root has been specified at or above AgrSP), but are missing T (which occurs below AgrSP). Because the root is specified as AgrSP or

above, the structure must project at least up to SA, including TP. Therefore whatever the reason for T omission in [-T] clauses, it cannot be due to truncation.

Furthermore, we must not dismiss the fact that Mood is obligatory in every utterance. Truncation makes no distinction between functional projections other than their respective hierarchical positions. Without the stipulation of a lower limit of what can be specified as the root, Truncation holds no potential to explain why Mood is obligatory and not any of the other functional affixes.

4.5.3 Underspecification of T

Wexler (1994) argued that in English and other RI languages children have the option to underspecify T. For Swahili, as we saw in section 4.2.6, the underspecification of T hypothesis predicts that child grammar permits full clauses and [-T] clauses. While these clause types are attested, they are by no means the only clause types that occur in child Swahili. In fact, at stage 1, full clauses and [-T] clauses account for only 39% of all indicative verbs, suggesting that the underspecification of T is only partly responsible for the typology of child clauses.

4.5.4 Underspecification of Agr

Clahsen *et al.* (1996) argued that AgrS is optionally underspecified. We saw that their theory predicts that Swahili children should allow full clauses and [-SA] clauses. However, they cannot account for the occurrence of [-T] clauses and bare stems. In fact, neither underspecification of T nor underspecification of Agr can account for bare stems.

4.5.5 ATOM

The natural conclusion from the preceding discussion is that the underspecification of T as well as Agr is permitted by child grammar – precisely the proposal of Schütze & Wexler (1996). They propose that both Agr and T can be optionally and independently omitted, resulting in the predictions that Swahili children should allow full clauses (no underspecification), [-SA] clauses (AgrS underspecification only), [-T] clauses (T underspecification only) and bare stems (underspecification of both AgrS and T). Thus our results are consistent with ATOM.

4.6 Remaining questions

While ATOM does predict the four clause types that occur in early Swahili, there are several aspects of the Swahili data which ATOM does not directly explain. We will now consider three results that ATOM does not address directly.

The first issue that ATOM does not account for is the difference between SA omission and T omission. We saw earlier that T omission is less frequent than SA omission at all times, and ceases to be a possibility well before SA omission ceases to be a possibility. ATOM does not predict this difference, but it is not incompatible with ATOM either. There is nothing in ATOM that either predicts this or rules this out.

The second issue is that of bare stems. According to ATOM, underspecification of Agr and T are independent. Therefore the omission of both SA and T (i.e., bare stems) should be the product of the independent omission SA and the independent omission of T. That is, the rate of bare stems should be related to the rates of [-SA] clauses and [-T] clauses since bare stems occur through the independent omission of SA and T. This is

clearly not the case in the data. In every stage except stage 4, bare stems outnumber [-T] clauses, which is unexpected if bare stems are a product of [-T] and [-SA] clauses.

However, Carson Schütze (p.c.) points out that this prediction holds only on individual child data and not on staged data. ATOM says nothing about aggregated data, and must be evaluated on a file-by-file basis. I include such data in Appendix 4C, and an inspection of this data reveal that bare stems in each file outnumber [-T] clauses. While this is not incompatible with ATOM, it strongly suggests that Agr omission and T omission are not entirely independent, and that there may be other factors that contribute to the omission of inflectional prefixes in Swahili.

The final issue relates to Mood. Deen & Hyams (2002) investigate the occurrence of subjunctive and indicative verbs and their interpretation. Their overall finding is that indicative verbs from early on occur in prototypically realis contexts while subjunctive verbs occur in prototypically irrealis contexts. However, they also found that in stages 1 and 2, Mood is not as productive as it is in stages 3 and 4. Not only are the overall frequencies lower in the earlier stages, but the type-token ratios are lower, and the proportion of subjunctive-to-indicative verbs is lower.

Table 4.12 Types/tokens of verbs expressing irrealis mood

	Stage 1	Stage 2	Stage 3	Stage 4
Type	8	5	26	20
Tokens	9	7	50	37

Table adapted from Deen & Hyams (2002)

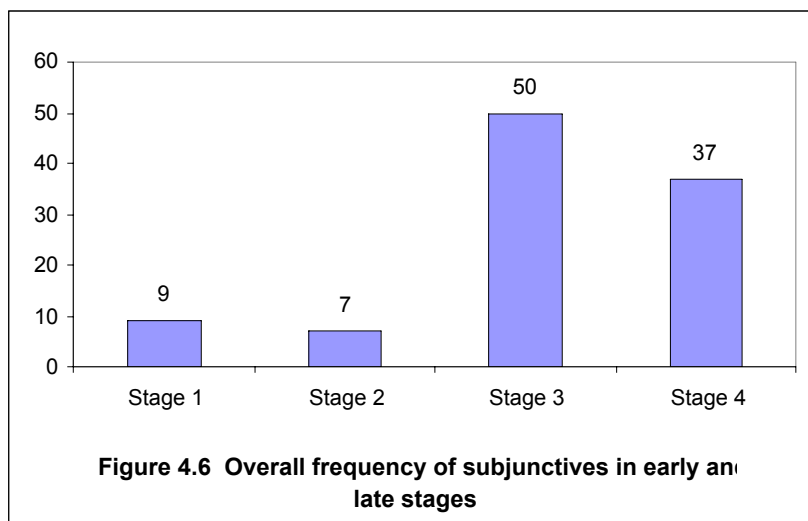


Table 4.13 Proportion of subjunctives to indicative verbs in early stages and late stages.

	Stages 1+2	Stages 3+4
Subjunctives	16	87
Indicatives	500	826
Ratio	0.03	0.10

Table 4.13 and figure 4.6 show the raw frequency of subjunctives in each stage. Table 4.13 then shows the ratios of subjunctives-to-indicative verbs in the early stages and the late stages. We see that in the early stages, subjunctive verbs occur at a rate of 0.03. This means that for every 100 verbs in these stages, 3 are subjunctive. In stages 3 and 4 the ratio jumps to 0.10, which means that proportionately there are three times as many subjunctive verbs in the later stages than the early stages. Deen & Hyams suggest that stage 3 is when Mood becomes productive.

Alternation in the final mood vowel is crucial in showing productivity in Mood in early grammar. The paucity of subjunctives in the

early stages, therefore, suggests that Mood too may be underspecified. Notice also that the stage in which Mood becomes productive (stage 3) co-occurs with the stage in which Tense becomes obligatory in the child grammar (see figure 4.4 in which we saw that T omission diminishes to around 10% by stage 3)⁹⁵.

Agr and T may be underspecified, and this results in omission of SA and T. If Mood is underspecified, then how then do we explain the earlier finding that the mood final vowel is virtually always present? There is good reason to believe that the phonotactic system of a particular language is acquired very early by children, perhaps before the first word is even spoken (e.g., Jusczyk *et al*, 1993). We saw in chapter 2 that Swahili generally does not allow coda consonants, but requires that every word end in a vowel. This is typically visible in loan words that take on a thematic final vowel, e.g., *lazima* (from the Arabic ‘ladhim’), *wiki* (from the English ‘Week’), etc. Children undoubtedly are attuned to this phonotactic requirement. Let us assume that Mood is underspecified. This would result in the child omitting the mood final vowel, just as SA is omitted through all four stages and T is omitted primarily through stages 1 and 2. Thus a target utterance such as (44a) with Mood underspecified would have an intermediate representation as in (44b). However, knowledge of the phonotactic requirement that coda consonants are ungrammatical would prevent (44b) from surfacing, and the form in (44c) would surface.

- (44) a. ni – ta–anguk–a
 SA_{1s}–fut–fall–IND
 ‘I will fall down.’

⁹⁵ See Deen & Hyams (2002) for some speculations about the connection between T and Mood.

- b. ni – ta – anguk
SA_{1s}–fut–fall–∅
- c. ni – ta – anguk – a
SA_{1s}–fut–fall–thematic vowel

However, as Carson Schütze points out, this predicts that the final vowel on the verb should be subject to the same principles that govern epenthetic vowels in loan words. The final epenthetic vowel in loan words is related to the preceding vowel in the stem, i.e., when the vowel in the stem is [high, front], the epenthetic vowel is harmonized as [high, front]. When the preceding vowel is [low, back], the epenthetic vowel is [low, back]. Thus the final vowel on the verb should alternate according to verb stems. This is clearly not the case, with the majority of verbs being marked with the indicative *a-*, including verbs that have [high, front] vowels in the stem, e.g., *ingi-a* ‘enter’, *pig-a* ‘hit’, etc. Thus the naturalistic data on this point are not conclusive, and this remains a point for future research.

4.7 Conclusion

In this chapter we investigated the acquisition of Swahili verbal inflection. We saw that overall Mood is rarely omitted, although we suggested that this may be due to phonotactic requirements. We saw that OA too is rarely omitted. We found that Swahili children omit SA and T in significant proportions at early stages, and slowly develop towards the adult norm. Swahili children independently omit both Agr as well as T. I showed that neither a metrical theory of omission nor a Truncation theory accounts for this. I also showed that underspecification of a single functional head is inadequate to account for all the data, and thus the

Swahili data lend support to Schütze & Wexler’s ATOM which states that Agr and T may be independently and optionally underspecified.

We will see in the next chapter that these inflectional omissions are not simply morphological or phonological omissions. While Swahili cannot be examined in the way that other European languages have been examined (because Swahili does not have wh- movement, null subjects are allowed, verb raising does not occur over negation or adverbs, there is no V2 effect, etc.), I will show that there are syntactic effects of this underspecification. I will show that different sorts of subject omission correlate with the different clause types, suggesting that the omission of inflectional prefixes is syntactic in nature.

Appendix 4A. Missing mood vowel examples

Below is an exhaustive list of the utterances in which the final vowel was omitted. In several cases, the meaning is unclear, but the best estimate has been provided. In some cases, the final vowel was missing as well as one or both prefixes. None of these examples were included in the calculation of the proportions of the various clause types discussed in chapters 4 and 5.

(1)	mama, ni – p ile ledio mama OA _{1s} –give–Ø that radio 'Mother, give me that radio.'	Has01, line 1202.	(8)	tak kojo – a Ø–Ø–want–Ø potty–IND 'I want to potty.'	Mus09, line 294
(2)	n –tak ...yyy MPH–want–Ø 'I want...'	Haw04, line 1435	(9)	tak koto – a Ø–Ø–want–Ø potty–IND 'I want to potty.'	Mus09, line 1098
(3)	i – zim apa ? MPH–shut off–Ø here 'Should I shut it off from here?'	Haw06, line 999	(10)	kap ... tini Ø–Ø–sit–Ø down 'I am sitting down.'	Mus14, line 109
(4)	hu – on nn sana neg–see–Ø much 'You don't see much.'	Haw07, line 130	(11)	si – m – pendm MaSaba neg–OA _{3s} –like–Ø MaSaba 'I don't like MaSaba.'	Fau04, line 1791
(5)	Ø – na – tak ku –ntm– pand – a Ø–pres–want–Ø inf–OA _{3s} –climb–IND 'I want to climb him.' ('I want to climb onto daddy..')	Mus12, line 2272	(12)	ne – sha – onn kumi SA _{1s} –p.comp.–see–Ø ten 'I have seen ten.'	Fau04, line 444
(6)	on dudu hapa see–Ø insect here 'Look at the insect here.'	Mus20, line 775			
(7)	ata ku– nfay hivo even inf–do–Ø this way 'You should do it this way.' (meaning unclear)	Mus21, line 605			

Appendix 4B. Root Infinitives

Below is an exhaustive list of the root infinitives that occur in the corpora. The infinitive marker in Swahili is *ku*, which occurs as a prefix. The infinitive prefix is homophonous with the 2nd person singular object agreement marker. Because of this, 9 additional examples have been omitted from this list as it was unclear whether they were RIs or OAs.

- | | | | |
|--|-----------------|---|------------------|
| <p>(1) ku – tap – a
inf–slap–IND
'(I will) slap him.'</p> | Mus03, line 26 | | |
| <p>(2) ku – tok – a hapa
inf–get out–IND here
'(I want) to get out of here.'</p> | Mus03, line 580 | | |
| <p>(3) ni – ku – sumbu – a
SA_{1s}–inf–disturb–IND
'(I will) disturb (her).'</p> | Mus08, line 11 | | |
| <p>(4) ku – tup – a
inf–throw–IND
'(I want) to throw (it).'</p> | Mus11, line 125 | | |
| <p>(5) ku – og – a
inf–bathe–IND
'(I want) to bathe.'</p> | Mus21, line 589 | | |
| <p>(6) mbaza ...aza ku – ni – chund – a
Mbaza inf–OA_{1s}–pinch–IND
'Mbaza will pinch me.'</p> | Mus22, line 833 | | |
| <p>(7) ku – tap – a mguu
inf–slap–IND foot
'(Daddy will) slap (my) foot.'</p> | Mus22, line 898 | | |
| <p>(8) na – tawadh – a wa ...we ...ku – tawadh – a
pres–bathe–IND you inf – bathe – IND
'I am bathing...you (should) bathe.'</p> | Has01, line 428 | | |
| | | <p>(9) ku – on – a wa-tu wa imbo
inf–see–IND 2–person 2-of song
'I see the person of the song.'</p> | Has01, line 1416 |
| | | <p>(10) ku – chez – a
inf–play–IND
'(I want / I'm going) to play.'</p> | Has01, line 1431 |
| | | <p>(11) Faiza na– ku – juan – a na, na, na hii chocolate?
Faiza pres–inf–know–IND and and and this chocolate
'Faiza knows (that this is) chocolate?'</p> | Has02, line 830 |
| | | <p>(12) na miskiti ku – ingi – a ta – kuj – a ...saa ngapi ?
and mosque inf–enter–IND fut–come–IND time what?
'And when you enter the mosque, what time will you come?'</p> | Has02, line 878 |
| | | <p>(13) ku – kul – a
inf–eat–IND
'(I want) to eat.'</p> | Has03, line 918 |
| | | <p>(14) ku – kul – a
inf–eat–IND
'(I want) to eat.'</p> | Has03, line 1266 |

None of these examples are acceptable in adult Swahili. There are reports of *ku-* used as a strong imperative (see chapter 2), but two of my consultants agree that these are not adult-like. Most of these examples involve a modal interpretation, either an expression of want or an intention. However, examples (9) and (11) are exceptions. Furthermore, in some cases, the *ku* could be interpreted as a misanalyzed subject agreement marker: because 2nd person singular OA is *ku*, the child may be using *ku* to mark second person singular in general. An example of this is (12), where the child uses a complex biclausal utterance. In the first clause the child uses a so-called

RI, and the second clause is a [-SA] clause. One interpretation of this is that the first clause is an RI. A second interpretation is that *ku* marks the second person subject. If *ku-* were indeed a misanalyzed SA marker, we might predict that in other utterances *ku-* occurs with a tense marker. However, there are no examples in the entire corpus of *ku-* followed by a tense marker.

Notice in examples (13) and (14), the child uses the infinitive marker with the verb *kula* 'to eat.' In Standard Kiswahili the verb 'to eat' is a monosyllabic verb root that takes dummy *ku* as a prefix. I argued in chapter 2 that in Nairobi Swahili *ku* has been reanalyzed as part of the verb stem because it is not omitted in any contexts. These examples provide additional evidence for this position since the child has analyzed the verb stem as *kula* and added an infinitive marker to it. If the child had a Standard Kiswahili analysis of the verb, the child would have produced *kula*, not the reduplicative *kukula*.

Appendix 4C. Monthly numbers of omissions by child

In chapter 3 I describe the methodology for pooling data into stages. I use three criteria to establish the four stages: MLU, V ratio, and the proportion of Proto-syntactic place holders. Based on this division of the data, I divide the following data into stages. The reason for this is so that a unified view of the data can be obtained. The complexity of having four children at different stages in development and of different ages makes this a difficult task. In the text I claim that the conclusions that I come to based on staged data are still valid when the data are arranged on a monthly basis. Thus below I provide the data arranged month-by-month.

Each table is for one child, with the ages in the left column and the clause types in each subsequent column. Importantly, there are very few cells in which nothing appears. In other words, all four of these children produced all four clause types at all times in their data recording. Therefore the omission of prefixes in child Swahili is not something limited to particular children, but rather is something that all children do – even the relative mature children.

Table 4C.1. Monthly data for Hawa

	Full Clauses	[-SA] Clauses	[-T] Clauses	Bare Stems	Total
2;2	5	1	0	2	8
2;3	2	6	2	12	22
2;4	1	5	5	19	30
2;5	3	3	5	8	19
2;6	2	5	6	19	32
Total	13	20	18	60	111

Table 4C.2. Monthly data for Mustafa

	Full Clauses	[-SA] Clauses	[-T] Clauses	Bare Stems	Total
2;0	1	0	1	2	4
2;1	2	3	0	0	5
2;2	5	23	14	16	58
2;3	18	14	9	2	43
2;4	33	71	16	18	138
2;5	14	43	4	10	71
2;6	1	6	0	4	11
2;7	5	17	4	13	39
2;8	5	17	1	10	33
2;9	9	16	3	9	37
2;10	43	15	1	8	67
Total	136	225	53	92	506

Table 4C.3. Monthly data for Fauzia

	Full Clauses	[-SA] Clauses	[-T] Clauses	Bare Stems	Total
1;8	6	12	1	4	23
1;9	34	44	3	4	85
1;10	4	20	2	1	27
1;11	33	21	8	6	68
2;0	64	12	1	1	78
2;1	42	26	2	1	71
Total	183	135	17	17	352

Table 4C.4. Monthly data for Hassan

	Full Clauses	[-SA] Clauses	[-T] Clauses	Bare Stems	Total
1;10	125	53	19	8	205
1;11	100	51	7	7	165
Total	225	104	26	15	370

Appendix 4D. Subjunctives and their interpretation

In the text I only include data for the omission of SA and T. Further analyses have been performed investigating the occurrence of subjunctive/indicative alternations, reported in Deen & Hyams, 2002. These data are presented here. Please refer to Deen & Hyams (2002) for a fuller discussion.

Table 4D.1 Total indicative, subjunctive and negative (final vowel) utterances by stage

Stage	Indicative	Subjunctive	Negative	Total
1	210	9	19	238
2	295	7	11	313
3	460	50	76	586
4	377	37	22	436

In table 4D.1 we see that indicative clauses are the most frequent clause types at all stages. This is also true of adult speech, and so this is unsurprising. However, the proportion of subjunctives increases by stage, as the figure and table below show:

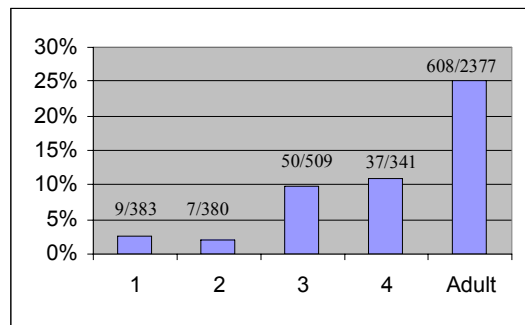


Figure 4D.1. Subjunctives as a proportion of non-imperative verbs

Table 4D.2 Subjunctives by stage

	Stage 1	Stage 2	Stage 3	Stage 4
Types	8	5	26	20
Tokens	9	7	50	37
Token/type ratio	0.88	0.71	0.52	0.54
Irrealis meanings	Desire Request Suggestion	Desire Request	Desire Possibility Request Permission Suggestion	Desire Possibility Request Permission Suggestion

Moving on to the interpretation of clauses, we find that a total of 23 out of 1436 indicative verbs are used incorrectly (indicated in bold in the table), i.e., 23 indicative verbs are used in a context that is most compatible with a subjunctive final vowel (predominantly a context in which the child is expressing his/her desire). This represents an error rate of 1.6%, which I consider to be negligible.

Table 4D.3 Interpretation of children's indicative verbs

	MEANING	
Adult Indicative	Present, on-going	737
	Present result	160
	Past	266
	Future	89
	Intentional	161
	Desire	17
Adult subjunctive	Possibility	0
	Necessity	0
	Request	0
	Suggestion	6
	Total	1436

The numbers in table 4D.3 above include all clause types, i.e., full clauses, [-SA] clauses, [-T] clauses and bare stems. It is possible that the presence of the tense marker in most of these indicative clauses forces an

indicative interpretation, and hence the low error rate that we see in table 4D.3 is a result of temporal marking. In order to control for that, I investigated the interpretation of bare stems, which are lacking tense marking. The results are presented below:

Table 4D.4 Interpretation of bare indicative stems

		MEANING	
Adult Indicative		Present, on-going	92
		Present result	18
		Past	25
		Future	6
		Intentional	18
		Desire	3
Adult Subjunctive		Possibility	0
		Necessity	0
		Request	1
		Suggestion	1
		Total	164

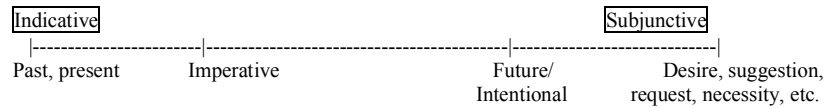
Here we see an error rate of 5/164 (the errors are indicated in bold), which is 3%. I still consider this a negligible rate of error, and so I conclude that children have knowledge of when the indicative final vowel is to be used. However, when we investigate the interpretation of subjunctive clauses, we find a different story. In table 4D.5, we see that 0f the 105 subjunctive clauses that occur in the entire corpus, a total of 18 are used in indicative contexts (errors are in bold). This is an error rate of 17%, something that can not be chalked up to speech error.

Table 4D.5 Interpretation of subjunctive verbs

		MEANING	
Adult Indicative		Present, on-going	2
		Present result	3
		Past	3
		Future	1*
		Intentional	9*
Adult Subjunctive		Desire	18
		Possibility	2
		Necessity	0
		Request	46
		Suggestion	21
		Total	105

Much cross-linguistic work has been done investigating the semantic contexts of subjunctive forms. Subjunctive is the morphology that is often associated with irrealis mood (Bybee, Perkins, and Pagliuca 1994; Chafe 1995; Givón 1994). Givón, for example, shows that irrealis morphology generally occurs with suggestions, the expression of desires, making of requests etc. Realis or temporal forms, on the other hand, occur in past and present contexts. Chafe (1995)describes these different semantic contexts as occurring in a “gradient” of markedness, as in figure 6, with the unmarked realis contexts and irrealis contexts on opposing ends of a continuum. In the unmarked case, imperatives pattern as realis forms and have indicative morphology, while future and intentionals pattern like irrealis forms and have subjunctive morphology. However, the morphology of these three categories – imperatives, futures and intentionals – is subject to cross-linguistic variation.

(1) The unmarked ‘gradient’ of markedness.



Swahili adheres to the unmarked realis-irrealis distinction with the exception of future and intentionals, which have indicative morphology and hence represent a marked option. Therefore, the Swahili system is a marked mood system in that it departs from the prototypical realis/irrealis mapping. The Swahili mood system is summarized in table 4D.6:

Table 4D.6: Summary of Swahili mood marking

Indicative/realis	Subjunctive/irrealis
Past	Desire
Present	Suggestion
Imperative	Necessity
Future	Request
Intentional	Permission
	Possibility

According to Chafe and Givón, adult Swahili future/intentional contexts are indicative and are hence marked (indicated by the dotted box above). The Swahili children, however, use the indicative-subjunctive morphology according to the unmarked mapping, extending subjunctive morphology to future and intentional contexts. The Swahili children have not acquired the marked characteristic of Swahili according to which future and intentionals have indicative morphology. If we put aside these 10 cases (marked in table 4D.5 by the asterisks) as representing a principled departure from the adult grammar, the number of subjunctive errors drops to 8 – that is, only 7.6%. The corrected form-meaning contingencies are presented in table 4D.7.

Table 4D.7 Form-meaning contingencies

	Indicative Marking	Subjunctive marking
Target Irrealis	5 (3.1%)	87 (83%)
Target Realis	159 (96.9%)	8 (7.6%)
	164	95

Our conclusion therefore is that Swahili children have not acquired the marked characteristic of the Swahili mood system, but show knowledge of mood distinctions at very early ages.

Chapter 5. Subject Omission and [-SA] clauses

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) | <ul style="list-style-type: none"> 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) | <ul style="list-style-type: none"> a. John is eating the apple English b. * is eating the apple | |
| | <ul style="list-style-type: none"> 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 languages⁹⁶. 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, 1999). P.Bloom (1990) argues for a processing solution that makes particular reference to VP length as the determining factor in processing load⁹⁷. 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

⁹⁶ An obvious exception is Valian (1991), see the discussion below.

⁹⁷ See Hyams & Wexler (1993) for arguments against processing accounts of subject omission in general, as well as specific criticisms of Bloom's (1990) proposal.

omission 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 subjects⁹⁸. 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

⁹⁸ 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 frequently 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.

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:

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 <i>is</i>
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- contexts⁹⁹.

- (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 picture 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

⁹⁹ 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 & Tuijnman). 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

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

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 occurring 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- contexts¹⁰⁰.

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:

$$(10) \text{ ForceP} > \text{TopP}^* > \text{FocP} > \text{TopP}^* > \text{FinP} > \text{IP}$$

(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

¹⁰⁰ 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%¹⁰¹. 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.

¹⁰¹ 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 time¹⁰². 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

¹⁰² 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.

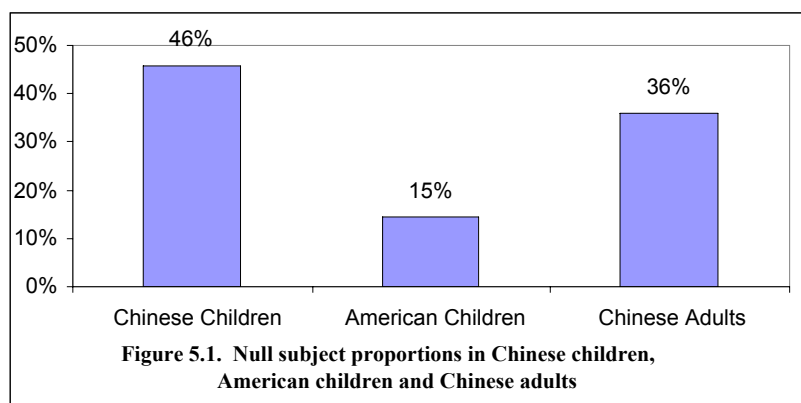


Table 5.12 Chinese and American Children’s proportions of null arguments

	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
this 3 rd inanim.- pr.prf.- spoil – IND
'This has spoiled' | Subject |
| (12) | huko alianguka
There 3 rd s – past – fall – IND
'He fell there' | Locative Adjunct |
| (13) | hile ninafikiri imeribika
That 1 st s-pres-think-IND 3 rd inanim.-pr.prf.-spoil-IND
'As for that thing, I think it is spoiled' | Topic |

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 required¹⁰³:

- (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:

¹⁰³ 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 contexts¹⁰⁴.

- (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’

¹⁰⁴ 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–OA₂–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

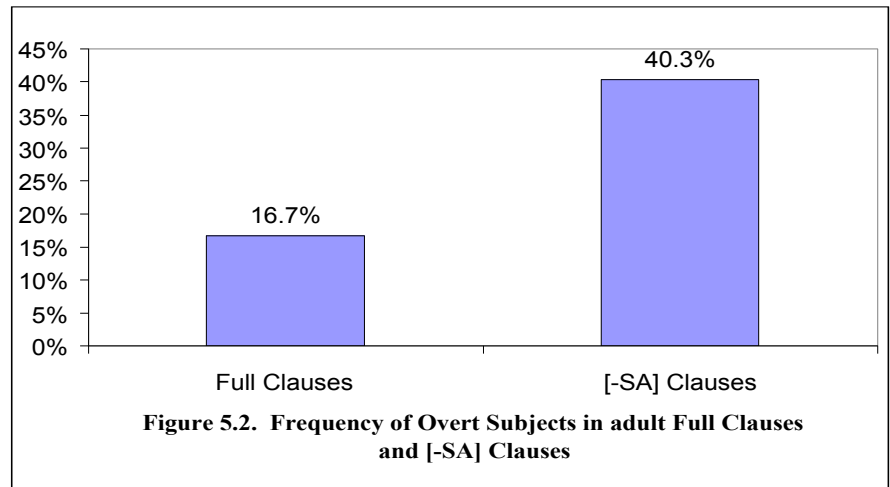
a raised voice) uses a bare stem. My judgment is that the bare verb is ungrammatical, but is acceptable in this context. My other native speaker consultant agrees. This is similar to the hypothetical English case in (31) in which the tense bearing element *do* is omitted:

- (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



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

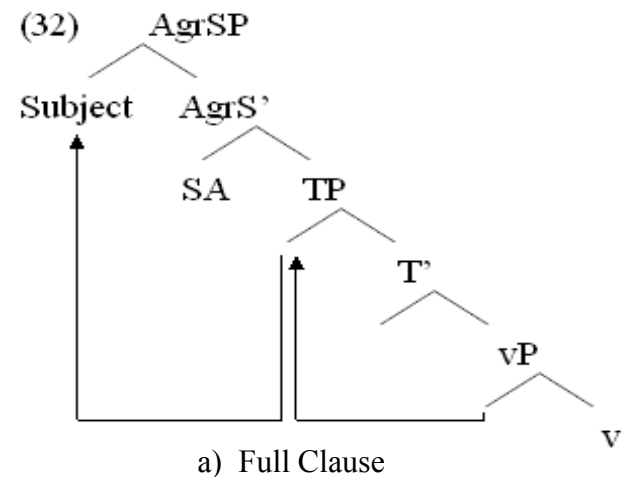
Recall that we found that overall, approximately 18% of verbal utterances contain overt subjects (see Table 5.13). However, we now know that there are at least two types of grammatical clauses in adult Nairobi Swahili: full clauses and [-SA] clauses. An automated CLAN analysis reveals that in full clauses, 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 (α -level = 0.01, $\chi^2 = 14.296$) according to a χ^2 test (with Yates Correction Factor).

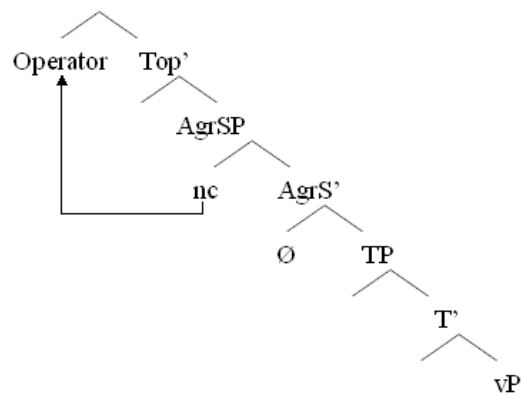
Thus the prediction that [-SA] clauses should only occur with overt subjects is false. Summarizing what we have found so far, overt subjects

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:





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

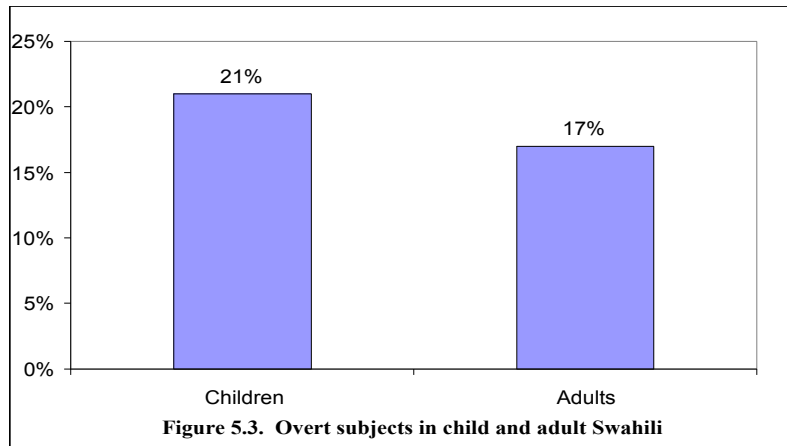
In looking at subjects in child Swahili, a useful place to start is to compare the rates of overt and null subjects to the adult rates. Using CLAN, a COMBO analysis of overt subjects in adult and child Swahili found that adults do not differ dramatically from children. Figure 3 presents aggregate data 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, subjects¹⁰⁵ occur in different proportions in full clauses and [-SA] clauses (17% and 40%, respectively).

¹⁰⁵ 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.



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.

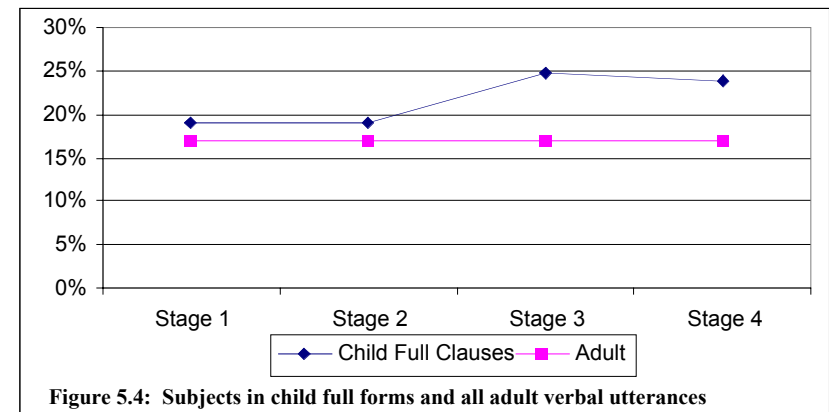
In the next four sections I will present results of analyses of subjects in each of the four clause types in child Swahili. I will conclude that lexical subjects and *pro* are permitted in full clauses only, while the null constant – topic operator construction is permitted in [-SA] clauses and bare 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
<i>Pro</i>	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 ($z=0.7921$).



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.

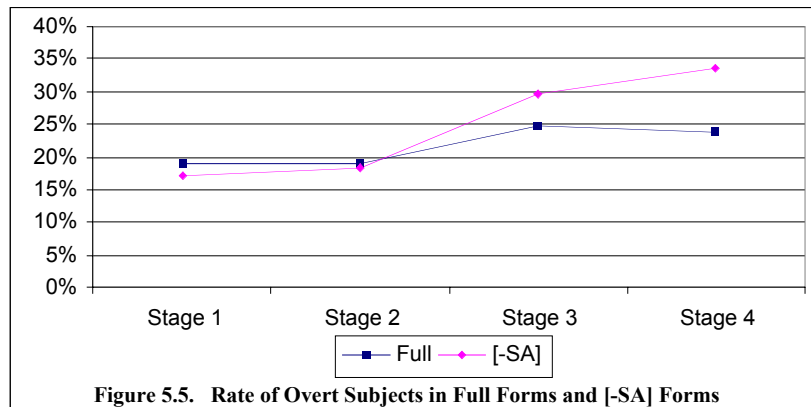


Figure 5.5. Rate of Overt Subjects in Full Forms and [-SA] Forms

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 SA₅ – pres – go – IND
 ‘The car is going’

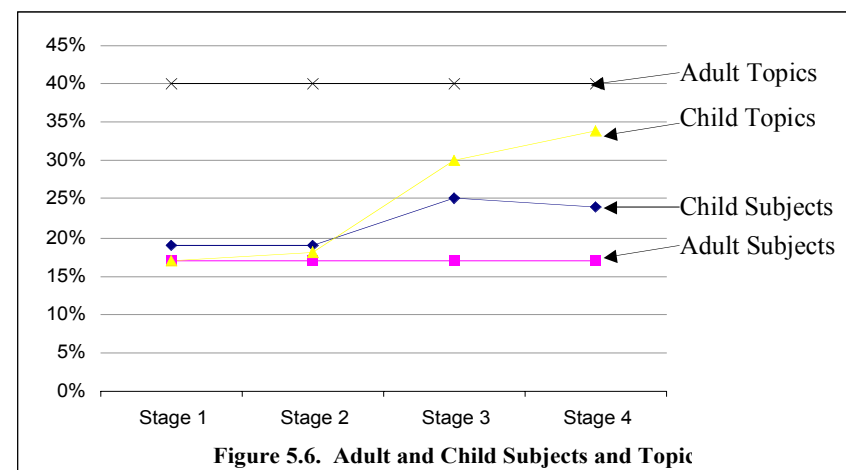
(40) Muko na Charlie mw–ingine na – va – a vi – atu FAU02 line 991
 Muko and Charlie 1-other pres–wear–IND 8–shoe
 Target = Muko na Charlie mw–ingine wa – na – va – a vi – atu
 Muko and Charlie 1-other SA_{3pl}–pres–wear–IND 8–shoe
 ‘Muko and the other Charlie are wearing shoes’

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.



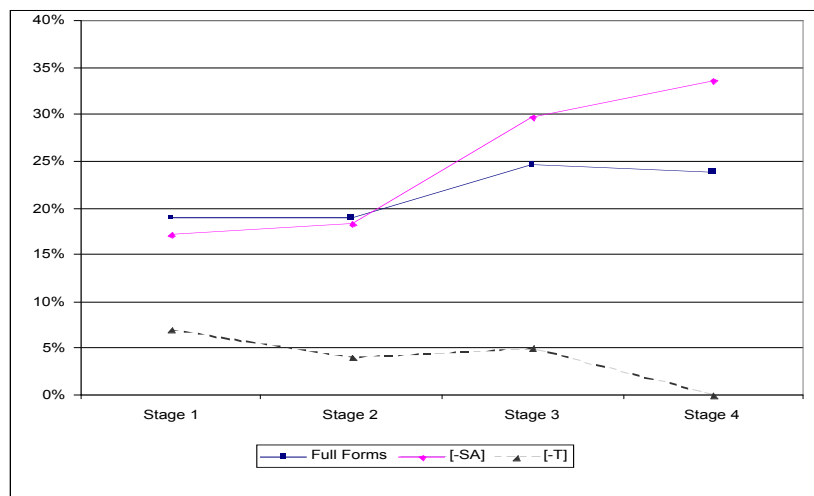
What is especially interesting is that [-SA] clauses are very rare in the adult language (approximately 5% of all sentences), so the properties of the different clause types either are acquired by children with the aid of very rare input or follow from general principles of grammar. In the next two sections 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 Φ -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 done 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 ($p=0.01$, $\chi^2=19.767$, with Yates Correction Factor), as is the difference between overt subject rates in [-T] clauses and [-SA] clauses ($p=0.01$, $\chi^2=23.293$, 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

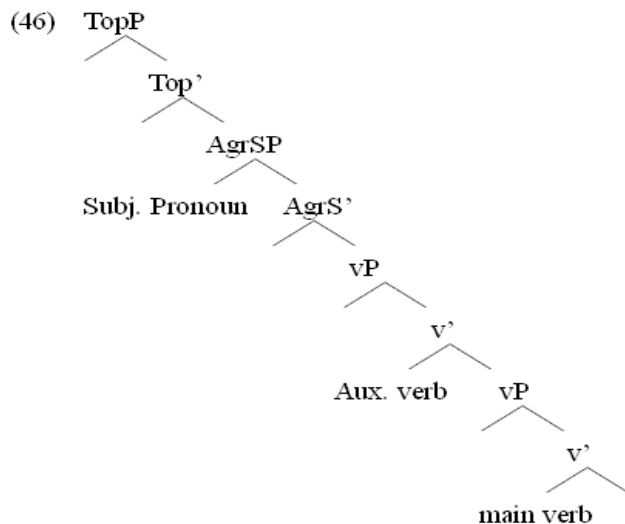
In fact, there were only 5 [-T] clauses in the entire corpus that had an overt subject. Below is the exhaustive list of all the [-T] clauses with subjects the children produced¹⁰⁶.

- (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
 ‘I am being quiet’

¹⁰⁶ 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.

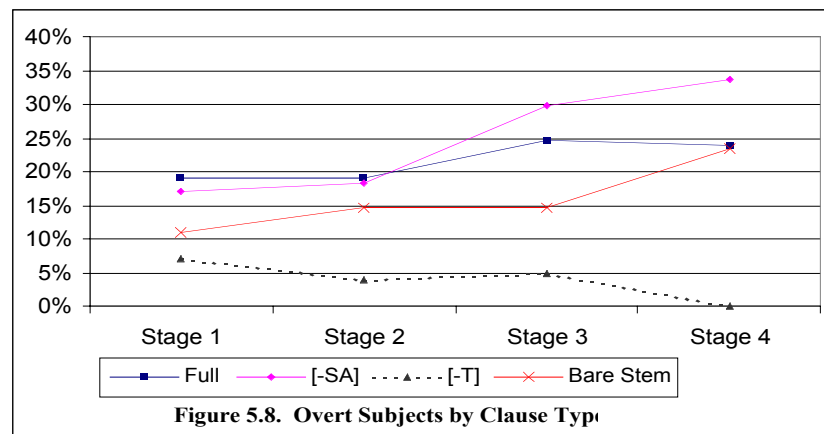
However, as I have just explained, these five examples are not entirely convincing. My point is simply that we should keep in mind that the count of 5/114 [-T] clauses with subjects is a generous one, and the actual number could well be lower than this.

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.



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 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 Φ -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 earlier¹⁰⁸.

¹⁰⁸ 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.

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) $\emptyset - \emptyset - V - \text{IND}$

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

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

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.

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 5.16¹⁰⁹, reproduced below.

Table 5.16 Summary of clause types and “subject” options

	Full Clause	[-SA] Clause	[-T] Clause	Bare Stem
<i>pro</i>	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 with the subject omission facts of Swahili.

¹⁰⁹ 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]]).

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 length¹¹⁰:

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

¹¹⁰ 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.¹¹¹

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–OA₇–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,

¹¹¹ 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.

suffixes, and 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 ($p=0.01$, $t=0.175$). 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.¹¹² 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).¹¹³ 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.

¹¹² 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.

¹¹³ 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	✓	✓
[-SA] Clause	✓	✓
[-T] Clause	*	✓
Bare Stem	✓	✓

Where '✓' 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	<i>pro</i>	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 <i>pro</i>	-
-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 intention 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.

References

- Aksu-Koç, Ayhan and Dan Slobin (1985). "The acquisition of Turkish." *The crosslinguistic study of language acquisition*. D. I. Slobin (Ed). New Jersey, Hillsdale Erlbaum. 1.
- Allen, George and Sarah Hawkins (1980). "Phonological Rhythm: Definition and development." *Child Phonology: Volume 1. Production*. G. Yeni-Komshian, J. Kavanagh and C. Ferguson (Eds). New York, Academic Press: 227-256.
- Archibald, John (1995). *Phonological Acquisition and Phonological Theory*. Hillsdale, NJ, Erlbaum Publishers.
- Ariel, Mira (1990). *Assessing NP antecedents*. London, Routledge.
- Ariel, Mira (1991). "The Function of accessibility in a theory of grammar." *Journal of Pragmatics* 16: 443-463.
- Ashton, Ethel Osteli (1947). *Swahili Grammar*. London, Longmans, Green and Co.
- Baker, Mark (1985). "The Mirror Principle and Morpho-syntactic Explanation." *Linguistic Inquiry* 16: 373-416.
- Baker, Mark (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago, University of Chicago Press.
- Barrett-Keach, Camilla (1986). "Word Internal Evidence from Swahili for AUX/INFL." *Linguistic Inquiry* 17, 3, 559-564.
- Bates, Elizabeth (1976). *Language and Context*. New York, Academic Press.
- Beghelli, Filipo and Tim Stowell (1992). "The direction of quantifier movement." Unpublished Manuscript. UCLA
- Behrens, Heike (1993). *Temporal Reference in German Child Language*. Amsterdam, University of Amsterdam.
- Blom, Elma and Frank Wijnen (2000). "How Dutch children's root infinitives become modal.." *Proceedings of the 24th annual Boston University Conference on Language Development*. S. Catherine Howell, Sarah A. Fish, and Thea Keith-Lucas (Eds). Somerville MA: Cascadilla Press, pp.128-139
- Blom, Elma, Evelian Krikhaar and Frank Wijnen (2001). "Root infinitives in Dutch and English child language: an experimental approach". *Proceedings of the 25th annual Boston University Conference on Language Development*. Anna H.-J. Do, Laura Domínguez, and Aimee Johansen (Eds) Somerville, MA: Cascadilla Press, pp. 133-144.
- Bloom, Lois (1970). *Language development: form and function in emerging grammars*. Cambridge MA, MIT Press.
- Bloom, Paul (1990). "Subjectless sentences in Child Language." *Linguistic Inquiry* 21(4, Fall): 491-504.
- Borer, Hagit and Ken Wexler (1987). The maturation of syntax. *Parameter Setting*. T. Roeper and E. Williams (Eds). Dordrecht, Reidel: 123-172.
- Bottari, Piero, Paola Cipriani and Anna Maria Chilosi (1993/1994). "Protosyntactic Devices in the Acquisition of Italian Free Morphology." *Language Acquisition* 3(4): 327-369.
- Börjars, Kersti and Carol Chapman (1998). "Agreement and pro-drop in some dialects of English." *Linguistics* 36(1): 71-98.
- Bresnan, Joan and Sam Mchombo (1987). "Topic, Pronoun and Agreement in Chichewa." *Language* 63: 741-782.
- Bromberg, Hilary Sarah and Kenneth Wexler (1995). "Null subjects in Child Wh-questions." *MIT Working papers in linguistics*, 26. Carson Schütze, Jennifer Ganger, and Kevin Broihier (Eds), pp. 221-247.
- Brown, Roger (1973). *A First Language: The Early Stages*. London, Allen and Unwin.
- Buell, Leston (2000). "Swahili Relative Clauses." Unpublished Master's Thesis, UCLA.
- Campos, Hector (1986). "Indefinite Object Drop." *Linguistic Inquiry* 17(2): 354-359.
- Cardinaletti, Anna (1997). "Subjects and Clause structure." *The new comparative syntax*. L. Haegeman (Ed). London, Longman: 33-63.
- Carstens, Vicki (1991). "The Morphology and Syntax of Determiner Phrases in Kiswahili." Unpublished Doctoral Dissertation, UCLA.
- Chimombo, Moira (1981). "Overgeneralization in negation: a comparison of first and second language acquisition." Unpublished Doctoral Dissertation, Columbia University.
- Chomsky, Noam (1986). *Barriers*. Cambridge, MA, MIT Press.
- Chomsky, Noam (1993). "The Minimalist Program." *The view from building 20*. K. Hale and S. Keyser (Eds). Cambridge, MIT Press.
- Chomsky, Noam (1995). *The Minimalist Program*. Cambridge MA, MIT Press.
- Cinque, Guglielmo (1999). *Adverbs and Functional heads*. New York, Oxford University Press.

- Clahsen, Harold and Martina Penke (1992). "The Acquisition of Agreement Morphology and its Syntactic Consequences: New Evidence on German child language from the Simone-Corpus." *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Acquisition*. J. Meisel (Ed). Dordrecht, Kluwer.
- Clahsen, Harold, Sonja Eisenbeiss and Martina Penke (1996). "Lexical learning in early syntactic development." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 129-160.
- Clancy, Patricia M. (1985). "The Acquisition of Japanese." *The Crosslinguistic study of Language Acquisition: The Data*. D. Slobin (Ed). Hillsdale, NJ, Lawrence Erlbaum Associates. **1**: 373-524.
- Cole, Peter (1987). "Null Objects in Universal Grammar." *Linguistic Inquiry* **18**(4): 597-612.
- Comrie, Bernard (1976). *Aspect: An Introduction to the study of Verbal Aspect and other related problems.*, Cambridge University Press.
- Connelly, Michael (1984). "Basotho children's acquisition of noun class morphology." Unpublished Doctoral Dissertation, University of Essex.
- Connelly, Michael (1987). "Basotho children's early development of speech." *African Studies* **46**(2): 229-239.
- Courtney, Ellen H. (1998). "Child acquisition of Quechua morphosyntax." Unpublished Doctoral Dissertation, University of Arizona.
- Crisma, Paula (1993). "On the acquisition of WH-questions in French." *GenGenP* **0**: 115-122.
- de Haan, Ger and K. Tuijman (1988). "Missing subjects and objects in child grammar." *Language Development*. P. Jordens and J. Lalleman (Eds). Dordrecht, Foris: 101-122.
- Deen, Kamil Ud (1997). "The interpretation of root infinitives in English: is eventivity a factor?" Term Paper, UCLA.
- Deen, Kamil Ud (1999). "Swahili verbal complexes: Morphology from Syntax." Qualifying Paper, UCLA.
- Deen, Kamil Ud (2001). "The Acquisition of Swahili verbal morphology". G.A.L.A., Palmela, Portugal.
- Deen, Kamil Ud and Nina Hyams (2002). "The Form and Interpretation of Finite and non-Finite verbs in Swahili". *Proceedings of the 26th annual Boston University Conference on Language Development*. Barbora Skarabela, Sarah Fish, and Anna H.-J. Do (Eds). Somerville, MA: Cascadilla Press, pp. 130-141.
- Deen, Kamil Ud (2002). "The omission of inflectional prefixes in the acquisition of Nairobi Swahili." Unpublished Doctoral Dissertation, UCLA
- Demuth, Katherine (1988). "Noun classes and agreements in Sesotho acquisition." *Agreements in Natural Language: Approaches, Theories and Descriptions*. M. Barlow and C. A. Ferguson (Eds), Center for the Study of Language and Information: University of Chicago Press: 305-321.
- Demuth, Katherine (1989). "Maturation and the acquisition of the Sesotho Passive." *Language* **65**: 56-80.
- Demuth, Katherine (1992). "Accessing functional categories in Sesotho: interactions at the morpho-syntax interface." *The Acquisition of Verb Placement: Functional categories and V2 phenomena in Language Acquisition*. J. Meisel (Ed). Dordrecht, Kluwer Academic Publishers: 83-107.
- Demuth, Katherine (1992). "The acquisition of Sesotho." *The cross-linguistic study of language acquisition*. D. Slobin (Ed). Hillsdale, NJ, Lawrence Erlbaum Associates. **3**.
- Demuth, Katherine (1993). "Issues in the acquisition of the Sesotho tonal system." *Journal of Child Language* **20**: 275-301.
- Demuth, Katherine (1994). "On the underspecification of functional categories in early grammars." *Syntactic Theory and first language acquisition: cross-linguistic perspectives*. B. Lust, M. Suner and J. Whitman (Eds). Hillsdale, NJ, Lawrence Erlbaum Associates.
- Demuth, Katherine (1995). "Questions, relatives and minimal projection." *Language Acquisition* **4**(1&2): 49-71.
- Demuth, Katherine and J. Gruber (1995). "Constraining XP sequences." *Niger-Congo Syntax and semantics* **6**. V. Manfredi and K. Reynolds (Eds). Boston, Boston University African Studies Center.
- Demuth, Katherine (1998). "Argument structure and the acquisition of Sesotho applicatives." *Linguistics* **36**(4): 781-806.
- Duffley, Patrick (1992). *The English Infinitive*. London, Longman.
- Duran, James (1975). "The Role of Swahili in a Multilingual Rural Community in Kenya." Unpublished Doctoral Dissertation, Stanford University.

- Echols, Catharine and Elissa Newport (1992). "The role of Stress and position in determining first words." *Language Acquisition* **2**: 189-220.
- Elbers, Loekie and Frank Wijnen (1992). "Effort, Production Skill, and Language Learning." *Phonological Development: Models, Research, Implications*. C. Ferguson, L. Menn and C. Stoel-Gammon (Eds). Timonium MD, York Press: 337-368.
- Enç, Murvet (1987). "Anchoring conditions for tense." *Linguistic Inquiry* **18**: 633-657.
- Ferdinand, Astrid (1996). *The development of Functional categories: The acquisition of the subject in French*. Dordrecht, ICG Printing.
- Feurer, H. (1980). "The Acquisition of Mohawk: Observations of First Language Development." *Studies in Iroquoian Culture*. N. Bonvillain (Ed). **6**: 119-130.
- Fikkert, Paula (1992). "The Acquisition of Dutch Stress". 17th Annual Boston University Conference on Language Development., Boston MA.
- Fletcher, Paul (1985). *A Child's learning of English*. Oxford, Basil Blackwell.
- Fox, Danny and Y. Grodzinsky (1998). "Children's passive: a view from the by-phrase." *Linguistic Inquiry* **29**: 311-332.
- Gerken, LouAnn (1991). "The metrical basis of children's subjectless sentences." *Journal of Memory and Languages* **30**: 431-451.
- Gerken, LouAnn and Bonnie McIntosh (1993). "The interplay of function morphemes and prosody in early language." *Developmental Psychology* **29**: 448-457.
- Givón, Talmy (1976). "Topic, pronoun and grammatical agreement." *Subject and Topic*. C. N. Li (Ed). London, Academic Press: 149-188.
- Grinstead, John (1998). "Subjects, Sentential Negation and Imperatives in Child Spanish and Catalan." Unpublished Doctoral Dissertation, UCLA.
- Guasti, Maria Teresa (1992). "Verb Syntax in Italian child grammar: finite and non-finite verbs." *Language Acquisition* **3**(1): 1-40.
- Guasti, Maria Teresa (1996). "Acquisition in Italian interrogatives." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 241-270.
- Guilfoyle, Eithne (1984). "The acquisition of tense and the emergence of lexical subjects in child English." *The McGill Working Papers in Linguistics* **2**(1): 20-30.
- Guthrie, Malcolm (1948). *The classification of the Bantu languages*. London, International African Institute.
- Gutman, Eynat (1999). "Null Subjects: A theory of syntactic and discourse-identification." Unpublished Doctoral Dissertation, University of Delaware.
- Haddon, E. (1955). *Swahili Lessons*. Cambridge, W. Heffer and Sons.
- Haegeman, Liliane (1994). "Root Infinitives, Tense and truncated structures." *Language Acquisition* **4**(3): 205-255.
- Haegeman, Liliane (1996). "Root infinitives, clitics and truncated structures." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 271- 308.
- Halle, Morris and Alec Marantz (1993). "Distributed Morphology and the pieces of inflection." *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. K. L. Hale and S. J. Keyser (Eds). Cambridge, MA, MIT Press.
- Hamann, Cornelia, Luigi Rizzi and Uli Frauenfelder (1996). "On the acquisition of subject and object clitics in French." *Generative Perspectives on Language Acquisition*. H. Clahsen (Ed). Amsterdam/Philadelphia, Benjamins.
- Hamann, Cornelia and Kim Plunkett (1998). "Subjectless sentences in Child Danish." *Cognition* **69**: 35-72.
- Harris, Tony and Kenneth Wexler (1996). "The optional-infinitive stage in child english: evidence from negation." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 1-42.
- Hayes, Bruce (1982). "Extrametricity and Stress." *Linguistic Inquiry* **13**: 227-276.
- Hinnebusch, Thomas J. (1979). "Swahili." *Languages and Their Status*. T. Shopen (Ed). Cambridge MA, Winthrop Publishers, Inc.

- Hoekstra, Teun and Nina Hyams (1995). "The syntax and interpretation of dropped categories in child language: a unified account". Chpt in *The Proceedings Of The Fourteenth West Coast Conference On Formal Linguistics*. Camacho, Jose, Choueiri, Lina, & Watanabe, Maki (Eds), Stanford, CA: Center Study Language & Information, 1996, pp 123-136.
- Hoekstra, Teun, Nina Hyams and Misha Becker (1996). "The underspecification of number and the licensing of root infinitives". *Proceedings of the 21st annual Boston University Conference on Language Development*. Elizabeth Hughes, Mary Hughes, and Annabel Greenhill (Eds). Somerville, MA: Cascadilla Press. pp. 293-306.
- Hoekstra, Teun and Nina Hyams (1998). "Aspects of Root Infinitives." *Lingua* **106**(1-4): 81-112.
- Huang, James C.-T. (1984). "On the distribution and reference of empty pronouns." *Linguistic Inquiry* **15**: 531-574.
- Huang, James C.T. (1989). "Pro-drop in Chinese: A Generalized control theory." *The Null Subject Parameter*. O. Jaeggli and K. Safir (Eds). Dordrecht, Kluwer Academic Publishers: 185-214.
- Hyams, Nina (1983). "Acquisition of parameterized grammars." Unpublished Doctoral Dissertation, CUNY.
- Hyams, Nina (1986). *Language acquisition and the theory of parameters*. Dordrecht, Reidel.
- Hyams, Nina (1987). "The Theory of Parameters and Syntactic Development." *Parameter Setting*. T. Roeper and E. Williams (Eds). Dordrecht, Reidel.
- Hyams, Nina and Kenneth Wexler (1993). "On the Grammatical Basis of Null Subjects." *Linguistic Inquiry* **24**: 421-459.
- Hyams, Nina (1996). "The underspecification of functional categories in early grammar." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 91-128.
- Ingram, David and William Thompson (1996). "Early syntactic acquisition in German: evidence for the modal hypothesis." *Language* **72**(1): 97-120.
- Jaeggli, Osvaldo and Nina Hyams (1988). "Morphological uniformity and the setting of the null subject parameter". *Proceedings to NELS 18*. J. Blevins & J. Carter (Eds).
- Jaeggli, Osvaldo and Kenneth Safir (1989). *The Null Subject Parameter*. Dordrecht, Kluwer.
- Jaeggli, Osvaldo and Kenneth Safir (1989). "The Null Subject Parameter and Parametric Theory." *The Null Subject Parameter*. O. Jaeggli and K. J. Safir (Eds). Dordrecht, Kluwer: 1-44.
- Jonas, D. (1995). "On the Acquisition of child syntax in child Faroese." MIT Working Papers in Linguistics **26**. Carson Schütze, Jennifer Ganger, and Kevin Broihier (Eds), pp. 265-280.
- Josefsson, G. (1999). "Non finite Root clauses in Swedish child language". *Proceedings of the 23rd annual Boston University Conference on Language Development*. Annabel Greenhill, Heather Littlefield, and Cheryl Tano (Eds). Somerville, MA: Cascadilla Press.
- Jusczyk, Peter W, Angela D Friederici, Jeanine M I Wessels, Vigdis Y Svenkerud and Ann Marie Jusczyk (1993). "Infants' Sensitivity to the Sound Patterns of Native Language Words." *Journal of Memory and Language* **32**(3): 402-420.
- Kayne, Richard (1994). *The Antisymmetry of Syntax*. Cambridge MA, MIT Press.
- Keach, Camillia (1995). "Subject and Object markers as agreement and pronoun incorporation in Swahili." *Theoretical approaches to African Linguistics*. A. Akinlabi (Ed). Trenton NJ, African World Press: 109-116.
- Khamisi, A. M. (1988). "A Typology of Gaps in Kiswahili sentences." *Kiswahili: Journal of the Institute of Kiswahili Research* **55**(1): 120-133.
- Kinyalolo, Kasangati (1991). "Syntactic Dependencies and the SPEC-Head Agreement Hypothesis in KiLega." Unpublished Doctoral Dissertation, UCLA.
- Koopman, Hilda and Dominique Sportiche (1991). "The position of subjects." *Lingua* **85**: 211-258.
- Koopman, Hilda and Anna Szabolci (2000). *Verbal Complexes*. Cambridge, MA, MIT Press.
- Krämer, Irene (1993). "The Licensing of Subjects in Early Child Language." *MITWPL* **19** (Papers on Case and Agreement II). Philips, Colin (Ed), pp. 197-212.
- Krifka, Manfred (1995). "Swahili." *Syntax: Ein Internationales Handbuch Zeitgenössischer Forschung. An International Handbook Of Contemporary Research*. J. Jacobs, A. von Stechow, W. Sternefeld and T. Vennemann (Eds). Berlin, Walter de Gruyter. **2**: 1397-1418.

- Kunene, Euphrasia (1979). "The Acquisition of SiSwati as a First Language: A Morphological Study with Special Reference to Noun Prefixes, Noun Classes and some Agreement Markers." Unpublished Doctoral Dissertation, UCLA.
- Kursawe, C. (1994). "Fragensätze in der Deutschen Kindersprache." Unpublished Master's thesis, University of Düsseldorf.
- La Belle, Christopher (1999). "A Longitudinal Study of Lexical and Prosodic Differentiation by a French/English Bilingual Child (1;6-2;4)." Unpublished Master's Thesis, UCLA.
- Lasnik, Howard and Tim Stowell (1991). "Weakest Crossover." *Linguistic Inquiry* 22: 687-720.
- MacWhinney, Brian. (2000). The CHILDES project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, Brian and Catherine Snow (1985). "The Child Language Data Exchange System." *Journal of Child Language* 12: 271-296.
- Maw, Joan and John Kelly (1975). *Intonation in Swahili*. Surrey, England, Unwin Brothers Limited.
- Meinhoff, C. (1932). *Introduction to the phonology of the Bantu languages*. Berlin, Dietrich Reimer/E. Vohsen.
- Meisel, Jürgen (1992). "Introduction: Functional Categories and Verb Placement in Language Development." *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Acquisition*. J. Meisel (Ed). Dordrecht, Kluwer.
- Meisel, Jürgen and Maria-Jose Ezeizabarrena (1996). "Subject-verb and object-verb agreement in early Basque." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 201-240.
- Mutaka, Ngessimo (1994). *The Lexical Tonology of Kinande*. Munchen, Germany, Lincom Europa.
- Myachina, Ekaterina Nikolaevna (1981). *The Swahili language: a descriptive grammar*. London, Routledge & Kegan Paul.
- Ngonyani, Deogratis (1996). "The Morphosyntax of Applicatives." Unpublished Doctoral dissertation, UCLA.
- Nurse, Derek and Thomas Hinnebusch (1993). *Swahili and Sabaki: A Linguistic History*. Berkeley, University of California Press.
- Park, Jae-Ick (1995). "Minimality Effects in Swahili." *Theoretical approaches to African Linguistics*. A. Akinlabi (Ed). Trenton NJ, African World Press.
- Peters, Anne (1997). "Strategies in the acquisition of Syntax." *The Handbook of child language*. P. Fletcher and B. MacWhinney (Eds). Oxford, Blackwell.
- Peters, Ann M. (2001). "Filler Syllables: What is their status in emerging grammar?" *Journal of Child Language* 28: 229-242.
- Phillips, Colin (1995). "Syntax at age 2: cross linguistic differences." *MIT Working Papers in Linguistics*. 26, Carson Schütze, Jennifer Ganger, and Kevin Broihier (Eds), pp. 325-382.
- Pierce, Amy (1989). "On the emergence of syntax: a cross linguistic study." Unpublished Doctoral Dissertation, MIT.
- Pizzuto, Elena and Maria Cristina Caselli (1991). "The acquisition of Italian morphology: implications for models of language development." *Journal of Child language* 19(3): 491-558.
- Platzack, Christer (1992). "Functional Categories and early Swedish." *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Acquisition*. J. Meisel (Ed). Dordrecht, Kluwer.
- Plunkett, Kim and Sven Strömquist (1990). "The acquisition of Scandinavian languages." *Gothenberg papers in theoretical linguistics* 59.
- Poeppel, David and Kenneth Wexler (1993). "The full competence hypothesis of clause structure in early German." *Language* 69: 1-33.
- Pollock, Jean-Yves (1989). "Verb Movement, UG and the structure of IP." *Linguistic Inquiry* 20: 365-424.
- Radford, Andrew (1986). "Small children's small clauses." *Bangor Research Papers in Linguistics*. 1: 1-38.
- Radford, Andrew (1990). "The Nature of children's initial grammars of English." *Logical issues in language acquisition*. I. M. Roca (Ed). Foris Publications: 199-235.
- Radford, Andrew (1990). *Syntactic theory and the acquisition of English Syntax*. Oxford, Basil Blackwell.
- Radford, Andrew (1996). "Towards a structure-building model of acquisition." *Generative perspectives on language acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. H. Clahsen (Ed). Philadelphia PA, John Benjamins Publishing Company: 43-90.
- Raghavendra, Parimala and Laurence Leonard (1989). "The acquisition of agglutinating languages: converging evidence from Tamil." *Journal of Child Language* 16: 313-322.

- Rasetti, Lucienne (2000). "Null subjects and root infinitives in the child grammar of French." *The Acquisition of Syntax*. M.-A. Friedemann and L. Rizzi (Eds). Longman: 236-268.
- Rizzi, Luigi (1982). *Issues in Italian Syntax*. Dordrecht, Foris.
- Rizzi, Luigi (1986). "Null objects in Italian and the theory of pro." *Linguistic Inquiry* 17: 501-557.
- Rizzi, Luigi (1992). "Early null subjects and root null subjects." *Language acquisition studies in generative grammar*. T. Hoekstra and B. Schwarz (Eds). Amsterdam/Philadelphia, John Benjamins: 151-176.
- Rizzi, Luigi (1994). "Some notes on Linguistic theory and language development: the case of root infinitives." *Language Acquisition* 3: 371-393.
- Rizzi, Luigi (1997). "The fine structure of the left periphery." *Elements of grammar. A Handbook of generative syntax*. L. Haegemann (Ed). Dordrecht, Kluwer.
- Rizzi, Luigi (2000). "Remarks on early null subjects." *The Acquisition of Syntax*. M.-A. Friedemann and L. Rizzi (Eds). Longman: 269-292.
- Roeper, Tom and Bernhard Rohrbacher (1994). "True pro-drop in child English and the principle of economy of projection." manuscript, UMASS
- Salustri, Manola and Julia Berger-Morales (2001). "Deficient Structures in Bilingual Children: Correlations between the Verbal Domain and Nominal Domain." Unpublished manuscript, UCLA.
- Salustri, Manola (2002). "Simultaneous Acquisition of German and Italian: A Longitudinal Study of Bilingual Children of Pre-School Age". *Proceedings of the 26th annual Boston University Conference on Language Development*. Barbora Skarabela, Sarah Fish, and Anna H.-J. Do (Eds). Somerville, MA: Cascadilla Press, pp. 572-583.
- Sano, Tetsuya and Nina Hyams (1994). "Agreement, finiteness, and development of null arguments." *NELS* 24. M. Gonzalez (Ed).
- Sano, Tetsuya (1995). "Roots in Language acquisition: A comparative study of Japanese and European languages." Unpublished Doctoral Dissertation, UCLA.
- Sano, Tetsuya (1997). "The comparative morpho-syntax of root infinitives in child languages". *The Proceedings Of The Fifteenth West Coast Conference On Formal Linguistics*, Agbayani, Brian, & Tang, Sze-Wing (Eds), Stanford, CA: Center Study Language & Information, pp 457-471.
- Santelmann, Lyn (1999) "The acquisition of determiners in child Swedish: interactions in prosodic and syntactic constraints." Talk given at GALA, Sept. 11, 1999.
- Sarma, Vajjyanthi (1999). "Case, Agreement and Word Order: Issues in the syntax and acquisition of Tamil." Unpublished Doctoral Dissertation, MIT.
- Saville-Troike, Muriel (1996). "Development of the Inflected Verb in Navajo Child Language. *Athabaskan Language Studies: Essays in Honor of Robert W. Young*. E. Jelinek, S. Midgette, K. Rice and L. Saxon (Eds). Albuquerque, University of New Mexico Press, pp. 137-192.
- Schaeffer, Jeanette (1990). "The Syntax of the subject in child language: Italian compared to Dutch." Unpublished MA thesis, University of Utrecht.
- Schieffelin, Bambi (1985). "The Acquisition of Kaluli." *The Crosslinguistic study of Language Acquisition: The Data*. D. Slobin (Ed). Hillsdale, NJ, Lawrence Erlbaum Associates. 1: 525-594.
- Schütze, Carson and Kenneth Wexler (1996). "Subject case licensing and English Root infinitives". Proceedings of the 20th annual Boston University Conference on Language Development. Andy Stringfellow, Dalia Cahana-Amitay, Elizabeth Hughes and Andrea Zukowski (Eds). Somerville, MA: Cascadilla Press, pp. 670-681.
- Schütze, Carson (1997). "INFL in child and adult language: Agreement, Case and licensing." Unpublished Doctoral Dissertation, MIT.
- Selkirk, Elisabeth (1986). *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge MA, MIT Press.
- Slobin, Dan (1973). "Cognitive prerequisites for the development of Grammar." *Studies of Child Language Development*. C. Ferguson and D. Slobin (Eds). New York, Holt, Rinehart and Winston.
- Slobin, Dan (1985). *Cross linguistic Study of language acquisition*. Hillsdale, NJ, Lawrence Erlbaum Associates.
- Smoczynska, Magdalena (1985). "The Acquisition of Polish." *The Crosslinguistic study of Language Acquisition: The Data*. D. Slobin (Ed). Hillsdale, NJ, Lawrence Erlbaum Associates. 1: 595-686.
- Stowell, Tim (1981). "Origins of phrase structure." Unpublished Doctoral dissertation, MIT.
- Suppes, Patrick (1973). "The semantics of children's language." *American Psychologist* 88: 103-114.

- Suzman, Susan (1982). "Strategies for acquiring Zulu concord." *South African Journal of African Languages* 2: 53-67.
- Suzman, Suzman (1991). "Language Acquisition in Zulu." Unpublished Doctoral Dissertation, University of Witwatersrand.
- Taraldsen, Knut Tarald (1978). *On the NIC, vacuous application and the that-trace filter*. Bloomington.
- Valian, Virginia (1991). "Syntactic subjects in the early speech of American and Italian children." *Cognition* 40, 21-81.
- Veneziano, Edy and Hermine Sinclair (2000). "The changing status of 'filler syllables' on the way to grammatical morphemes." *Journal of Child Language* 27: 461-500.
- Vitale, Anthony (1982). "Problems of Stress Placement in Swahili." *Studies in African Linguistics* 13(3): 325-330.
- Vitale, Anthony (1985). "Aspects of Kiswahili Stress." *Kiswahili: Journal of the Institute of Kiswahili Research* 52: 106-113.
- Vitali, Anthony (1981). *Swahili Syntax*. Dordrecht, Foris Publications.
- Vollmann, Ralf (1999). "A case of 'morphotactic underspecification': The many functions of the fillers in an Austrian child." Available online at: <http://www-ang.kfunigraz.ac.at/~vollmanr/epub/VR1999E.html>.
- Webelhuth, Gert (1992). *Principles and parameters of syntactic saturation*. New York, Oxford University Press.
- Weissenborn, Jürgen (1990). "Functional Categories and Verb Movement: the acquisition of German Syntax reconsidered." *Sprachewerb und Grammatik. Linguistische Untersuchungen zum Erwerb von Syntax und Morphologie*. M. Rothweiler (Ed). Goettingen, Linguistische Berichte, Sonderheft 3.
- Weissenborn, Jürgen (1992). "Null Subjects in Early Grammars: Implications for Parameter Setting Theory." *Theoretical Issues in Language Acquisition*. J. Weissenborn, H. Goodluck and T. Roeper (Eds). Hillsdale, NJ, Lawrence Erlbaum.
- Weverink, Meike (1989). "The subject in relation to inflection in child language." Unpublished MA thesis, University of Utrecht.
- Wexler, Kenneth (1994). "Optional Infinitives, head movement, and economy of derivation." *Verb Movement*. N. Hornstein and D. Lightfoot (Eds). Cambridge, UK, Cambridge University Press: 305-350.
- Wexler, Kenneth (1998). "Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage." *Lingua* 106: 23-79.
- Wijnen, Frank (1996). "Temporal reference and eventivity in root infinitives." *MIT Occasional papers in Linguistics* (12). Jeannette Schaeffer (Ed), pp.1-25.
- Zanuttini, Raffaella (1997). *A Comparative Study of Romance Languages*. New York, Oxford University Press.
- Ziesler, Yasmine L. and Katherine Demuth (1995). "Noun class prefixes in Sesotho child-directed speech". 26th Annual Child Language Research Forum, Stanford, CSLI.
- Zwart, Jan Wouter (1997). "Rethinking Subject Agreement in Swahili." Manuscript, NOW/University of Groningen.