Subject Clitics and Subject Extraction in Somali

Andrew Paul Hubbertz

M.A. Thesis

Department of Linguistics
McGill University, Montreal

July 1991

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Arts

Abstract

This study applies Chomsky's Barriers version of government binding theory to subject clitic pronouns and subject-verb agreement processes in Somali. It is proposed that Somali has a contrast between strong and weak subject-verb agreement, indicated as AGR_S and AGR_W respectively. SCL is possible iff $AGR=AGR_S$. SCL is obligatory if subject is pro, the null pronominal. SCL is excluded in case of short subject extraction, which is only possible from the domain of AGR_W. The prohibition against short subject extraction from the domain of AGR_S is attributed to Principle B of Aoun's generalized binding theory. Long subject extraction from the domain of AGR_S is possible if an intermediate trace in [NP,CP] deletes before binding theory. The analysis is extended to subject extraction in Italian, including the Trentino dialect, and Modern Irish.

Résumé

Dans notre étude nous nous proposons d'appliquer la théorie de gouvernement et de liage, telle qu'elle paraît dans la version Barriers de Chomsky, aux prénoms clitiques sujet (SCL) et aux processus en somali de l'accord sujet-verbe. Nous pensons que le somali présente des contrastes entre l'accord fort (AGR_S) et faible (AGR_W) du sujet-verbe. Le SCL est obligatoire si le sujet est pro, cela veut dire que le prénom est vide. Le SCL est exclu en cas de l'extraction courte du sujet laquelle n'est possible que du domaine de AGR_W. On attribue au Principe B de la théorie généralisée du liage, mise sur le compte de J. Aoun, l'interdiction de l'extraction courte du sujet du domaine de AGR_S. Par contre, l'extraction longue du sujet du domaine de AGR_S est possible si on abandonne pro sur le terrain de l'extraction et qu'on efface la trace intermédiaire dans [NP,CP] avant d'appliquer la théorie du liage. Notre analyse de l'extraction du sujet comprend l'italian, surtout le dialecte du Trentino, et l'irlandais moderne.

Contents

Acknowledgements

I wish to thank my professors and fellow students in the McGill University Department of Linguistics for their stimulating company. I am particularly indebted to Mark Baker for advice and encouragement. Thanks are also due to the University of Saskatchewan for leaves of absence, which permitted me to undertake this project; in particular, I thank Nancy Brown and Paul Wiens, past and present Directors of Libraries at the University of Saskatchewan, for their support. Finally, I wish to thank my Somali consultants for their patient assistance.

1. Introduction

1.1 Objectives and Theoretical Framework

The objective of this study is to extend the Barriers version of government binding theory to a range of phenomena in the syntax of Somali, a Cushitic language spoken in Somalia, Djibouti, and adjacent parts of Ethiopia and Kenya. There are essentially two reasons for undertaking a study of this kind. In the first place, given the nativist claims of generative grammar, one expects to find common principles underlying the diversity of human languages. The claim is that certain principles derived from the close study of English, Romance, and other relatively well-studied languages, are attributable to the human speech faculty. As these principles are part of our biological endowment, we expect to find them at work in other languages as well. Therefore, extending the theory to Somali may permit an account of the known facts of the language that achieves a higher degree of explanatory adequacy. To the extent that such a project is successful, it also supports the claims made for the theory.

At the same time, the problems that are inevitably encountered in such an undertaking may contribute to the development of the theory. In recent years, the extension of transformational generative grammar to 'new' languages has been an especially fruitful area of research. Classic examples include Jaeggli (1982), for discussion of clitics and null objects in Romance, Rizzi (1982, c.2), for comparative treatment of Subjacency in English and Italian, and Kayne (1975 and subsequent work) for extensive and varied work on French.

The present study will examine a range of phenomena in Somali syntax which have been attested since the earliest

western literature on the language. The basic facts are well documented, and I have relied in large part on the published literature, as described in section 1.2 below, supplemented by approximately thirty-six hours of work with native speakers. In the course of this study, I shall move from questions that are particular to Somali, to questions of greater generality and theoretical interest. It will be shown that Somali has a contrast between strong and weak subject-verb agreement that accounts for a wide range of syntactic phenomena, and that Somali is, in the relevant sense, a null subject language. Finally, I shall propose a new analysis of subject extraction in null subject languages, based upon the Barriers framework, and incorporating the generalized binding theory of Aoun (1985).

This study assumes the government binding theory of Chomsky (1981), and the most recent version of this theory, as developed in Chomsky (1986, henceforth Barriers), works cited therein, and more recent work within the same framework. Chomsky (1981) is usually associated with 'government binding theory', and Barriers with the 'Barriers framework' or 'approach'. These terms are convenient, although they may be misleading, insofar as they suggest more discontinuity between these versions of the theory than in fact exists.

Familiarity with the theory is assumed here, and the reader is referred to the works cited for discussion and exposition of the same. Nevertheless, it will be useful to sketch the broad outline of the theory and to define a number of terms, for the sake of reference later in this study, and to minimize confusion that might arise because some terms have had varying usage, as the theory has developed.

The leading idea behind Barriers is that certain structures are 'barriers' to government and bounding, and should be given a unified treatment, such that government is blocked by a single barrier, while bounding, which is subject to a weaker locality condition, is blocked by two barriers.

Hence, we have a definition of 'barrier' (2), based in part upon the definition of a blocking category (1):

(1) Blocking category (Barriers, 14)

> c is a blocking category (BC) for b iff c is not L-marked (i.e. not directly theta-marked by a lexical category) and c dominates b.

(2) Barrier (Barriers, 14)

> c is a barrier for b iff (a) or (b)
> (a) c immediately dominates d, d a BC for b;
> (b) c is a BC for b, c ≠ IP

c is understood (Barriers, 14) to be a maximal projection, and 'immediately dominate' is understood to be a relation between maximal projections.

As well, Chomsky proposes a Minimality Condition, which has the effect of reducing ambiguity of government. The basic idea is that a cannot govern b if there is a 'closer' category which governs b, even if a would otherwise govern b:

(3) Minimality Condition (Barriers, 42)

> c is a barrier for b if c is the immediate projection of d, a zero-level category distinct from b.

Note that Chomsky also considers a wider definition of the Minimality Condition in which c is a barrier if it is 'a' projection. The narrower definition in (3) permits a specifier to be governed from outside a maximal projection, while the broader definition does not.

Given this definition of barrier, we may proceed to the definition of 'government' (5), itself based in part on the definition of 'm-command'[1] (4):

(4) M-command (Barriers, 8)

a m-commands b iff a does not dominate b, and every maximal projection that dominates a dominates b.

(5) Government (Barriers, 8)

a governs b iff a m-commands b and there is no c, c a barrier for b, such that c excludes b.[2]

In a typical case like (6), a will not govern b if c is a barrier for b:

(6) ...a...[$_c$...b...]

By the Minimality Condition, in (7) c is a barrier for b if c is the immediate projection of d, a zero-level category distinct from b (Barriers, 42); in this configuration, a cannot govern b:

(7) ...a...[$_c$...d...b...]

We shall also need to define proper government, a stronger relationship than government alone:

(8) Proper government (Barriers, 17)

a properly governs b iff a theta-governs or antecedent governs b.

Theta-government may be defined as in (9):

(9) Theta-government

a theta-governs b iff a is a zero-level category that theta-marks b, and a, b are sisters.

Theta-government may be viewed as feature sharing between a head and the theta-grid of a maximal projection that it theta-marks, along the lines of Stowell (1981). The core case is proper government by a verb of its complement. Antecedent

government holds between the links of a chain, and is expressed by means of co-indexing.

Proper government plays a prominent role in the theory, in particular in conjunction with the Empty Category Principle (ECP):

(10) Empty Category Principle (Barriers, 16; Lasnik and Saito, 257)

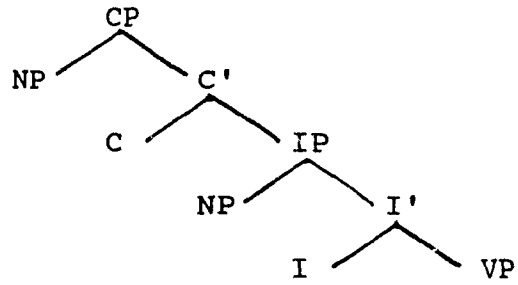
Trace must be properly governed.

Originally, the ECP applied to all empty categories (Chomsky 1981, 250). The Barriers framework, following Lasnik and Saito, restricts the ECP to traces.

Government, proper government, and the ECP have been part of the theory at least since Chomsky (1981); the idea of barriers and the Minimality Condition were introduced in Barriers, though, as always, one may find antecedents in the earlier literature.

To the above, must be added the following element of the Barriers framework which shall figure in this study. Barriers presents a revised version of X-bar theory, in which IP (= S) and CP (= S') are projections of I ('Inflection') and C ('Complementizer') respectively. IP and CP, to use the current terminology, are therefore not defective structures within the X-bar system. This has a number of consequences, including c-command of the subject ([NP,IP]) from the specifier of CP. English phrase structure, therefore, is essentially as in (11):

(11)



These brief notes will suffice for the present; I shall return to them as necessary.

1.2 Somali Orthography and Related Issues

In 1972, the Somali Democratic Republic adopted a standardized orthography for the Somali language, based upon the northern dialect, and using a western alphabet. Livnat (1984, 5-6) describes the system as phonetic, with most symbols representing 'standard phonetic transcription'. The principle exceptions, as noted in Livnat (1984) and Saeed (1987), are the following:

c	voiced pharyngeal fricative
x	voiceless pharyngeal fricative
sh	voiceless palatal fricative
j	voiceless palatal affricate
kh	voiceless velar fricative
'	glottal stop
dh	voiced retroflex alveolar stop
q	voiceless uvular stop

As well, two consecutive vowels represent a long vowel. See Saeed (1987, 13-26) for more details. Livnat (1984) and Saeed (1987) use standard Somali orthography. Examples taken from sources using other transcriptions will be converted to standard orthography without comment.

Hyman (1981) characterizes Somali as a tonal accent language, with tone introduced by morphological categories,

features, and construction types. The following are samples of the use of tone:[3]

(12) Natural gender (Hyman 1981, 172)

masculine		feminine	
náil	'lamb (m)'	naíl	'lamb (f)'
ínan	'son, boy'	inán	'daughter, girl'
nácas	'stupid man'	nacás	'stupid woman'
qaálin	'young camel (m)'	qaalín	'young camel (f)'

(13) Number (Hyman 1981, 172)

singular	plural	
túug	tuúg	'thief'
kálax	kaláx	'ladle'
bálli	ballí	'water reservoir'
Soomaáli	Soomaalí	'Somali man/Somali people

(14) Nominative Case (Saeed 1987, 20)

nominative[4]	non-nominative	
shabeelkii	shabeelkíi	'the leopard'
nin	nín	'man'

The Case distinction in (14) will arise in section 2 of this study. However, for the most part, tone is not relevant to the range of phenomena discussed here. Tone is not indicated in standard Somali orthography.

The fact that Somali has an official orthography suggests that the language is more standardized than is actually the case. There seems to be significant dialectal variation, although Somalis claim that all dialects are mutually intelligible. This problem is not entirely surmountable at present. In order to place this work on the solidest possible foundation, I have restricted my discussion to a core of widely documented phenomena respecting focus, subject clitics, subject-verb agreement, and nominative Case marking. Although

there have been various accounts of these phenomena, all studies agree on the basic facts.

At this point it is appropriate to acknowledge my debts to previous work on Somali. Although I had approximately thirty-six hours of field work with Somalis living in Montreal and Saskatoon, Canada, I owe a great deal to the work of B.W. Andrzejewski, F. Antinucci, C.R.V. Bell, C. El-Solami-Mewis, L. Hyman, M.A. Livnat, A. Puglielli, J.I. Saeed, and F. Serzisko. I have particularly relied upon Livnat (1984), whose approach to the issues treated here is closest to my own.

Notes

1. 'M-command' is based upon the earlier notion of 'c-command', which is still current, although it will not be used here:

i. c-command (Reinhart 1983, 18)

a c-commands b iff the branching note immediately dominating a also dominates b.

2. The question of 'exclusion' is peripheral to the main argument here. Chomsky proposes (Barriers, 6) that adjunction is possible only to a maximal projection that is a nonargument. One such case is adjunction of a complement of V to VP, creating a structure like (i):

(i) $[_b$ a $[_b$ ...]]

In this adjunction structure, b has two segments. Following May (1985), Chomsky proposes (Barriers, 7f.) that a category is dominated by b only if it is dominated by every segment of b. Consider then an adjuction structure like (ii):

(ii) ...d...$[_c$ a $[_c$...b...]]

Suppose c is a barrier. Then in order to insure that c governs b, we shall say that d 'excludes' a, but c does not. In particular,

(iii) a excludes b if no segment of a dominates b.

3. Saeed (1987) uses acute and grave accent to indicate high and falling tone, respectively; low tone is unmarked. Hyman (1981) reduces the Somali tone system to high and low tone. I have selected samples from Hyman (1981) that are consistent with Saeed (1987).

4. Saeed (1987) uses the terms 'subject' and 'absolutive' Case.

2. Somali Syntax and Subject Clitic Pronouns

In the first part of this section, I shall introduce some basic facts concerning Somali syntax. It is not my intent to provide a general introduction to the language; for that, the reader is referred to Saeed (1987), El-Solami-Mewis (1987), Bell (1953), and the authors cited at the end of part 1. Only those features will be discussed that are relevant to the issues treated here.

After this basic orientation, I shall examine the series of 'weak' or 'pre-verbal' subject pronouns in Somali, which will be shown to be subject clitic pronouns (SCLs). This discussion will be essentially theory-neutral, and is preliminary to section 3, where SCLs, agreement processes, and null subjects will be discussed within a government binding framework.

2.1 Introduction

Let us begin by considering a couple of simple sentences like (1). Note that SVO and SOV word-order are both quite common; they are presented here without prejudice.

(1)

a. nin-kii wuu arkay naag-tii
man-the F-he saw woman-the
'The man saw the woman'

b. nin-kii naag-tíi wuu arkay
(same meaning)

<u>-kíi</u>/<u>-tíi</u> are respectively masculine and feminine determiners. In standard Somali, these determiners display Case[1] through tone: non-nominative Case has high tone, which is lowered under nominative Case assignment. High tone is

also lowered in sentence-final position (Livnat 1984, 28, n. 2), so that naagtii has lowered tone in (1)a. There is a second set of determiners in which Case is marked by vowel alternation: -ka/-ta for non-nominative, and -ku/-tu for nominative. The latter series is generally used for proximate reference, while -kíi/-tíi is used for remote reference, i.e. with respect to events that are distant or in the past. The distribution of these determiners is summarized in (2):

(2)

	masc.		fem.	
	+nom.	-nom.	+nom.	-nom.
proximate	-ku	-ka	-tu	-ta
remote	-kii	-kíi	-tii	-tíi

No attempt will be made here to formulate the morphological rules underlying this paradigm.

The verb arkay displays subject agreement for person, number, and gender. Somali verbs exhibit rich inflectional morphology, including an alternation between 'extensive' and 'restrictive' paradigms, reflecting relatively strong and weak subject-verb agreement.[2] The restrictive paradigm is used in the case of short subject extraction by WH-movement, focus, or relativization; the extensive paradigm occurs elsewhere, including long subject extraction from sentential complement. The extensive and restrictive paradigms for the past tense of the verb 'to see' are illustrated in (3):[3]

(3)

	extensive	restrictive
1s	arkay	arkay
2s	aragtay	arkay
3sm	arkay	arkay
3sf	aragtay	aragtay
1p	aragnay	aragnay
2p	aragteen	arkay
3p	arkeen	arkay

The extensive paradigm does not have distinctive forms for every person-number; nevertheless, it reflects richer subject-verb agreement than the restrictive paradigm.

wuu represents the verb focus particle (FP) waa plus the 3sm subject clitic (SCL) uu. Every indicative matrix clause, and only the matrix clause, has one, and only one, FP. In the case of (1), this is the verb FP waa, and occurs in the outer layer of a complex of pre-verbal particles.[4] SCLs co-occur with the extensive paradigm only, and SCL and extensive paradigm conjointly distinguish all person-numbers, as in (4). Observe that neither SCL nor verb alone can uniformly distinguish the person-number of the subject.

(4)

1s	aan	arkay
2s	aad	aragtay
3sm	uu	arkay
3sf	ay	aragtay
1p(inc)	aynu	aragnay
1p(exc)	aanu	aragnay
2p	aydu	aragteen
3p	ay	arkeen

waa was described above as the verb FP, although in fact it is unclear whether it focuses V, VP, or whether it is neutral with respect to focus. Livnat (1984, 95-99) takes it to be a verb focus marker', while Saeed (1984, 184f.) rejects this approach. NP focus is, in this respect, less problematic. The focused NP moves to sentence-initial position and is followed by the FP baa or ayaa.[5] (5) illustrates NP focus in a simple sentence. Note that the FP baa + the 3sm SCL uu becomes buu, comparable to wuu above. NP focus movement is particularly clear in (5)b, where the focused object and baa or buu appear before the subject.

(5)

a. ninkíi baa/*buu arkay naagtii
man-the F /*F-he saw woman-the
'<u>The man</u> saw the woman'

b. naagtíi baa/buu ninkii arkay
woman-the F /F-he man-the saw
'The man saw <u>the woman</u>'

(5)a and (5)b have the same meaning, apart from focus.[6] Note that the 3sm SCL <u>uu</u> is optional in (5)b, but is ungrammatical in (5)a. Accounting for this distribution will be one objective of this study.

In (6)a below, the subject of the sentential complement has been extracted to the matrix clause, where it appears in sentence-initial position, followed by <u>baan</u> , representing the FP <u>baa</u> and the matrix clause SCL <u>aan</u>. In the case of subject extraction from a sentential complement, the SCL and the extensive verb paradigm are required in the sentential complement, as in (6)a. In (6)a and (6)b, the SCL in the sentential complement has cliticized onto the complementizer <u>in</u>.

(6) Livnat (1984, 43)

a. Cali baan mooday [inuu Berbera ka yimid]
Ali F-I thought that-he Berbera from came
'I thought that <u>Ali</u> came from Berbera'

b. Maryan baan mooday [inuu Cali arkay]
Maryan F-I thought that-he Ali saw
'I thought Ali saw <u>Maryan</u>'

Any NP may be focused, including adjuncts and NP extracted from a sentential complement.

What I am calling NP focus, with the FP <u>baa</u> or <u>ayaa</u>, may also be used to focus a clause, as in (7):

(7) Saeed (1987, 236)

a. inuu sugo ayaan u sheegay
that-he wait F-I to told
'I told him <u>to wait</u>'

b. inaan tago ayaan doonayaa
that-I go F-I want
'I want <u>to go</u>'

The choice of constitutent to focus is determined in part by discourse constraints, so that (9)a, but not (9)b or (9)c, is an appropriate response to (8). Note that (9)b and (9)c are fully grammatical, but are inappropriate in the given discourse.

(8) yaa arkay naagtii
who saw woman-the
'Who saw the woman?'

(9)

a. ninkíi baa arkay naagtii[7]
man-the F saw woman-the
'<u>The man</u> saw the woman'

b. #naagtíi buu ninkii arkay
woman-the F-he man-the saw
'The man saw <u>the woman</u>'

c. #ninkii wuu arkay naagtii
man-the F-he saw woman-the
'The man <u>saw</u> the woman' (or '<u>saw the woman</u>')

I shall not deal with the pragmatic and discourse functions determining focus. I shall assume, however, that the generation of FPs and movement (if any) of FPs and focused NPs occurs in the syntactic component of the grammar.

At the outset, I stated that SVO and SOV word order, as in (1)a and (1)b, are both quite common. In fact, Somali has relatively free word order, so that (1)a and (1)b are synonymous with each other and with (10)a to (10)c, and (11)a to (11)c are also synonymous with each other:

(1)

a.	ninkii wuu arkay naagtii man-the F-he saw woman-the 'The man saw the woman'	S F-V O
b.	ninkii naagtíi wuu arkay	S O F-V

(10)

a.	naagtíi ninkii wuu arkay	O S F-V
b.	naagtíi wuu arkay ninkii	O F-V S
c.	wuu arkay ninkii naagtii	F-V S O

(11)

a.	naagtíi buu ninkii arkay woman-the F-he man-the saw	O-F S V
b.	ninkii naagtíi buu arkay	S O-F V
c.	naagtíi buu arkay ninkii	O-F V S

The variations under (1), (10), and (11) are not equally valued, and some (e.g. (10)) are quite marked. Aside from the effects of NP focus movement, SVO and SOV are by far the most common word orders. See Livnat (1984, 7) for a similar range of data. Although some of the above are marked structures and are rather unnatural, they are of a different order of acceptability from (12)a and (12)b, which are entirely ungrammatical.

(12)

a.	*ninkii wuu naagtíi arkay	S F O V
b.	*arkay ninkíi baa naagtii	V S-F O

Expressed in terms that are appropriate to this stage of the discussion, (12)a is ungrammatical because a NP occurs between the verb focus complex <u>wuu</u> and the verb. (12)b is

ungrammatical because a focused NP occurs in post-verbal position.

It is misleading to suggest that Somali has free word order, insofar as that is understood to mean that constituents are freely 'scrambled'. Consider (10)c, in which the subject is in post-verbal position. This is only possible if the SCL uu is present (recall, waa+uu --> wuu); there is characteristically a pause before the post-verbal subject (cf. Saeed (1984, 36)); and the sentence is appropriate only within contexts determined by rather ill-understood discourse constraints. This suggests that Somali word order is not at all free, but that there are features of the language that permit a range of NP movement not available in a language like English. Nonetheless, the variation exemplified by (1), (10), and (11) complicates the task of determining basic, i.e. D-structure, word order in Somali.

Finally, it is appropriate to indicate that there are several common types of Somali sentence that will not be investigated here, as in (13), (14), and (15).

(13)

a. walaalkay waa barre
brother-my F teacher
'My brother is a teacher'

b. tareenkii waa raagay
train-the F late
'The train is late'

c. ninka guursaday walaalshay waa alwaaxqorre
man-the married sister-my F carpenter
'The man who married my sister is a carpenter'

(14)

a. duqdii waxay bilisay (waa) biyo
old woman-the what-she boiled F water
'The old woman boiled water'

b. Axmed wuxuu cabay (waa) sigaar
Ahmed what-he smoked F cigarette
'Ahmed smoked a cigarette'

(15) Livnat (1984, 91-92)

a. waxaa lacagta rabay ninka
F money-the wanted man-the
'The man wanted the money'

b. wuxuu yaqaan ninka dheeri Maryan
F-he knows man-the tall Maryan
'The tall man knows Maryan'

These sentences display verbless predication. (13)a to (13)c are of the structure 'NP waa X', where X is predicated of NP. The sentences in (14) and (15) employ waxaa, the so-called 'heralding' focus;[8] as in (13), the final NP is focused.

2.2 Subject Clitics

I have been claiming, in effect, that the subject pronouns in such sentences as (1), (5), and (6) above are subject clitics (SCLs). It is now time to examine that claim explicitly. Somali has two series of subject pronouns, as in (16):[9]

(16)

	SCL	full pronoun
1s	aan	aniga
2s	aad	adiga
3sm	uu	isaga
3sf	ay	iyada
1p(inc)	aynu	innaga
1p(exc)	aanu	annaga
2p	aydu	idinka
3p	ay	iyaga

The standard literature on Somali employs a variety of terminology with respect to what I am calling 'subject

clitics' and 'full' pronouns. Bell (1953) calls them 'simple' and 'emphatic' pronouns. Andrzejewski (1964) calls them 'preverbal' and 'substantive' pronouns. Antinucci and Puglielli (1980 and 1984) call them 'short' and 'long' forms. Serzisko (1984) calls them 'abhaengig' pronouns and 'unabhaengig' or 'emphatisch' pronouns. Saeed (1984 and 1987) calls them 'weak' or 'verbal' pronouns in the case of the former, 'independent', 'substantive', or 'emphatic' in the latter. El-Solami-Mewis (1987) calls them 'Kurzformen' and 'emphatische Formen'. Livnat (1984) calls them 'clitic' and 'full' pronouns. Livnat also considers the clitics to be resumptive pronouns when they occur in conjunction with subject extraction from a sentential complement, as in (6)a. I have adopted the terms 'clitic' and 'full' pronoun from Livnat, although my analysis of their status and location is substantially different. Underlying the terminological confusion is a variety of assumptions as to the syntactic status of what I am calling SCLs. Saeed (1984), for example, evidently assumes that 'weak' pronouns occur in argument positions, a point on which I shall differ. However, all of the above authors agree on the basic facts, and in particular on the distribution of these pronouns.

For expository reasons, I shall temporarily use the term 'weak' subject pronoun for what I have been calling 'subject clitics'. Given the uncertainty as to the status of 'weak' subject pronouns, it is appropriate to review evidence in support of the claim that they are, in fact, subject clitics.

Zwicky (1985) and Kayne (1974) offer a number of diagnostics for distinguishing clitics from full pronouns. If they apply to Somali weak subject pronouns, then these pattern with the more familiar cases of clitics, particularly from the Romance languages. It must be understood that these diagnostics represent generalizations rather than sets of rules. Exceptions may be found to many of them. In fact, it is unclear whether 'clitic pronoun' is even a valid category,

and future research may lead us to abandon the term entirely.[10]

Zwicky (1985) offers a number of diagnostics for distinguishing clitics from independent words, expressed in terms that are relatively theory-neutral. In particular, these diagnostics focus on the affix-like character of clitics. Not all of these diagnostics can be applied readily, in part for lack of information. However, the following criteria offer relatively clear results.

Zwicky (1985) observes that an element that is strictly ordered with respect to adjacent morphemes is a clitic or an affix, not an independent word. The weak subject pronouns in Somali are restricted in distribution to two positions (though I shall argue in section 4 that they are generated from a single position): clitized onto the NP FP <u>baa</u> or <u>ayaa</u>, in which case the FP + weak subject pronoun occur immediately after the focused NP; or cliticized onto the verb FP <u>waa</u>, in which case the weak subject pronoun occupies a determinate position in the outer layer of pre-verbal morphemes. These are illustrated in (17) and (18) respectively.[11] The weak pronouns are underlined.

(17) NP focus

a. askariga b<u>ay</u> gabadhu aragtaa
soldier-the F-she girl sees
'The girl sees the soldier'

b. guri ay<u>ay</u> fundiyaalku dhisayaan
house F-they craftsmen-the build
'The craftsmen are building a house'

(18) Verb focus

a. gabadhu w<u>ay</u> aragtaa askariga
girl-the F-she sees soldier-the
'The girl sees the soldier'

b. Maxamed wuu imanaya maanta
Mohammed F-he arrive today
'Mohammed will arrive today'

They may also appear cliticized onto a WH-word, as in (19):

(19)

a. muxuu Cali keeni doonaa
what-he Ali bring will
'What will Ali bring?'

b. Livnat (1984, 109)

yuu arkay Cali
who-he saw Ali
'Who did Ali see?'

There are two constructions that are problematic for this criterion, in that the weak subject pronoun appears, or can appear, as an independent word. The first of these is a relative clause, when it is a non-subject that is relativized. In this context, a weak pronoun may appear, either alone, or 'doubling' a lexical NP subject. Note that the FP is not permitted in a subordinate clause.

(20) Relative clause

a. Saeed (1987, 229)

buugga aan kuu keenayo
book-the I to you bring
'the book which I am bringing to you'

b. Livnat (1984, 56)

miiska uu Cali saaray buugga
table-the he Ali put book-the
'the table on which Ali put the book'

As well, a weak subject pronoun may appear in a sentential complement as an independent word. In most cases,

it cliticizes onto the complementizer in, but cliticization is not obligatory, as in (21)b:

(21)

a. Axmed baa u malaynaya [inuu berito booqanayo
Ahmed F thinks that-he tomorrow visit

barraha]
teacher-the
'Ahmed thinks that he will visit the teacher tomorrow'

b. Axmed baa u malaynaya [in berito uu booqanayo
Ahmed F thinks that tomorrow he visit

barraha]
teacher-the
(same meaning)

In subject extraction from a sentential complement, a weak subject pronoun is present near the extraction site, i.e. after the complementizer in and before the verb. The pronoun is co-referential with the extracted NP:

(22)

a. yuu$_i$ u maleenayaa Cali [inuu$_i$ Faarax arkay]
who to thought Ali that-he Farah saw
'Who did Ali think saw Farah?'

b. ardayda$_i$ ayuu barruhu faray [inay$_i$
students-the F-he teacher-the made that-they

buugga akhriyan
book-the read
'The teacher made the students read the book'

Zwicky (1985) also observes that clitics are never morphologically complex. That is true of the weak pronouns, which are invariant. Full pronouns, however, have the proximate determiner (cf. (16) above) and, in fact, may be inflected for Case, so that non-nominative aniga becomes nominative anigu.[12] This evidence also supports the claim that weak pronouns are SCLs.

An independent word is a syntactic constituent and can be subject to syntactic processes; a clitic is part of a word-like complex. In 2.1, I indicated that Somali has relatively free word order. NP movement is equally available to lexical NP and full pronoun. However, such movement is unavailable to weak pronouns. In (23) and (24), a lexical NP subject and a full pronoun respectively have been moved to post-verbal position. Such movement is impossible for weak subject pronouns, whether or not the focus particle is moved with it, as in (25) and (26):

(23)

wuu iibsaday Cali faras
F-he bought Ali horse
'Ali bought a horse'

(24)

wuu iibsaday isagu faras
F-he bought he horse
'He bought a horse'

(25)

*waa iibsaday uu faras
F bought he horse
'He bought a horse'

(26)

*iibsaday wuu faras
bought F-he horse
'He bought a horse'

Similarly, NP focus may apply to a full pronoun, as in (27), as freely as to a lexical NP, but cannot apply to a weak subject pronoun.

(27)

aniga baa libaax arkay
I F lion saw
'I saw a lion'

Kayne (1974, 81f.) includes an extended discussion of clitic pronouns in French, and in that context provides a number of criteria for distinguishing clitics from full lexical pronouns. I shall review only those diagnostics not also discussed in Zwicky (1985).

Kayne (1974, 84) observes with respect to French, that nothing may intervene between a SCL and the verb, as in (28):

(28)

a. *Il, paraît-il, est fou.
'He, it appears, is crazy.'

b. Jean, paraît-il, est fou.'
'Jean, it appears, is crazy.'

For Somali, this generalization is true of weak subject pronouns occuring with <u>waa</u> (cf. (18)), but it is false of those occuring with <u>baa</u> or <u>ayaa</u> (cf. (17)) or with WH-words (cf. (19)). It is patently not true of weak subject pronouns in relative clauses (cf. (20)) or in sentential complements (cf. (21) and (22)). This apparent anomaly must be set aside for now. What certainly can be observed at this point is that French SCLs and Somali weak subject pronouns are alike in enjoying restricted distribution. In particular, they cannot occur in isolation, as may be seen in (29) and (30), where B are grammatical, and C are ungrammatical responses to A.

(29)

A Qui part?
'Who is leaving?'

B lui, moi
him, me

C *il, *je
he, I

(30)

A: waa kuma ardayga ugu maskaxdafiican fasalka?
Q who student-the to-there smartest class-the
'Who is the smartest student in the class?'

B: isaga, iyada, aniga
him, her, me

C: *uu, *ay, *aan
he, she, I

Kayne (1974, 86-87) notes that for many speakers of French, full pronouns ('strong' pronouns in his terminology) are not readily used in reference to inanimates. Hence, (31)a is doubtful, because the full pronoun <u>lui</u> refers to an inanimate. (31)b is entirely acceptable, because it is the clitic <u>le</u> that refers to the inanimate.

(31)

a. ?Ce livre-lá, elle ne lit plus que lui.
'That book, she no longer reads anything but it.'

b. On le lit partout, ton bouquin.
'People are reading it everywhere, your book.'

Jaeggli (1982, 41) makes the same observation with respect to Spanish. (32)a and (32)b are ungrammatical, because the full pronoun <u>élla</u> refers to an inanimate; (32)c, in which the clitic <u>la</u> refers to the inanimate, is grammatical.

(32)

a. *La mesa, vimos (a) élla en esa tienda.
b. *La mesa, la vimos (a) élla en esa tienda.
c. La mesa, la vimos en esa tienda.

'The table, we saw it in that store.'

The same facts obtain in Somali, some speakers evidently further restricting full pronouns to [+human]. Observe the contrast between (33) and (34).

(33)

a. gurigii wuu gubtay
house-the F-he burned
'The house burned'

b. wuu gubtay
F-he burned
'It (the house) burned'

c. *isagu wuu gubtay
he F-he burned
'It (the house) burned'

(34)

a. faraskii wuu turanturooday
horse-the F-he tripped
'The horse tripped'

b. wuu turanturooday
F-he tripped
'It (the horse) tripped'

c. isagu wuu turanturooday
he F-he tripped
'It (the horse) tripped'

Kayne (1974, 90) further observes that clitics do not occur with modifiers, e.g.:

(35)

a. *Partiront-ils deux?
b. *Viendrez-vous autres?

The same facts are easily confirmed of weak pronouns in Somali. The full pronoun <u>iyaga</u> can be modified, as in (36)b, but the weak pronoun in (36)c cannot.

(36)

a. beeralayda oo dhan baa abuuraya galay maanta
farmers-the all F plant maize today
'All the farmers are planting maize today'

b. iyaga oo dhan baa abuuraya galay maanta
they all F plant maize today
'All of them are planting maize today'

c. *ay oo dhan baa abuuraya galay maanta
they all F plant maize today
'All of them are planting maize today'

As well, Kayne (1974, 90) notes that clitics cannot be conjoined, as in (37). The same may be confirmed of weak pronouns in Somali in (38):

(37)

a. *Partiront il et elle?
b. *Viendrez tu et Jean?

(38)

a. Cali iyo Axmed baa abuuraya galay
Ali and Ahmed F plant maize
'Ali and Ahmed are planting maize'

b. Cali iyo isaga baa abuuraya galay
Ali and him F plant maize
'Ali and he are planting maize'

c. *Cali iyo uu baa abuuraya galay
Ali and he F plant maize
'Ali and he are planting maize'

d. *uu iyo Cali baa abuuraya galay
he and Ali F plant maize
'He and Ali are planting maize'

Interestingly, with respect to one range of facts observed by Kayne (1974, 87-88), French patterns quite differently than Somali. In French, clitics are inflected for Case, as in (39); full pronouns are not.

(39)

a. Ils sont partis.
'They left.'

b. Cet enfant les voit
'That child sees them.'

c. Sa mère leur parlera
'His mother will speak to them.'

As noted above, Somali full pronouns are morphologically complex and may be inflected for Case. Weak subject pronouns differ from weak object pronouns, but they also occur in different positions. Therefore, they must be regarded as distinct constitutents, rather than as a single set which is inflected for Case. Note as well that the set of weak object pronouns is defective, lacking a third person, which makes them relatively more affix-like.[13] Be that as it may, it is clear from the evidence that Case is of no use as a diagnostic for distinguishing clitic and full pronouns in Somali.

To the above, I would propose an additional generalization: if full and clitic pronouns are both available for a particular Case, the full pronoun must be 'clitic doubled', i.e. if the full pronoun occurs, the clitic must occur as well. Jaeggli (1982, 40-41) notes that this is true for Spanish:

(40)

a. *Vimos a él

b. *Encontramos a élla

c. *Juan visitó a mi ayer

(41)

a. Lo vimos a él
'We saw him'

b. La encontramos a élla
'We found her'

c. Juan me visitó a mi ayer
'Juan visited me yesterday'

Compare these with (42):

(42)

a. Vimos a Pedro
'We saw Pedro'

b. Encontramos a Marie en el parque
'We found Marie in the park'

c. Juan visitó a sus padres ayer
'Juan visited his parents yesterday'

Romanian has both full and clitic accusative and dative pronouns (Murrell and Ştefănescu-Drăgăneşti, 1970). The full pronoun is always clitic doubled in the accusative, and 'usually' clitic doubled in the dative:[14]

(43)

a. Accusative (M & S, 168)

l-am	văzut	pe el	ieri
him-I have	seen	him	yesterday

'I saw him yesterday'

b. Dative (M & S, 131)

ce le place lor să facă?
what them pleases them to do
'What do they like to do?'

In fact, full subject pronouns in Somali require the presence of the weak pronoun (if available), suggesting therefore that the latter is a SCL, as in the following:

(44)

a. isagu wuu /*waa arkay
he F-he F saw
'He saw (him/her/it/them)'

b. Livnat (1984, 24)

ninkíi baan/*baa anigu arkay
man-the F-I F I saw
'I saw the man'

Note that there are two contexts in Somali which are exceptions to the generalization. If a full subject pronoun is focused, it appears in non-nominative Case, and the weak subject pronoun is not permitted, as in (45)a. If a full pronoun occurs in the 'NP <u>waa</u> X' construction (described p. 16-17 above), the full pronoun occurs in nominative Case and without weak subject pronoun, as in (45)b.

(45)

a. isaga baa arkay
he(-nom.) F saw
'He saw (him/her/it)'

b. El-Solami-Mewis (1987, 55)

isagu waa arday
he F student
'He is a student'

The considerations above have been directed to evidence that weak subject pronouns in Somali are in fact subject

clitics, rather than independent words. However, as is well known, clitics occupy an indeterminate area between affixes and fully independent words. Having argued that they are not fully independent words, it is equally important to establish that they are not affixes. In fact, the behavior of weak pronouns in subordinate clauses ((20), (21), and (22)) is strong evidence that they are not affixes, particularly in sentential complements, where they may appear anywhere between the verb and the complementizer. If anything, these weak pronouns are more 'independent' than the typical clitic, and hence more word-like. Nevertheless, Zwicky and Pullum (1983, henceforth Z & P) offer a number of diagnostics for distinguishing clitics from affixes. Without wishing to belabour the point, I shall briefly review some of their criteria and demonstrate that these weak pronouns are not affixes.

Z & P (p. 503) say that clitics have a low degree of selection with respect to hosts, while affixes have a high degree of selection with respect to stems. Saeed (1987, 27f.) identifies three categories of verb in Somali: weak verbs, strong verbs, and the irregular verb <u>yahay</u> ('to be'), where these categories are broadly comparable to their equivalents in European languages. He also (Saeed 1987, 120-7) categorizes nouns into seven declensions. The weak subject pronouns in Somali show no selection with respect to any of these verb or noun categories. Moreover, the possible hosts for weak subject pronouns are restricted to FPs, WH-words, and the complementizer <u>in</u>. These comprise closed classes of words and therefore should not normally take affixes;[15] hence the weak subject pronouns attached to them are presumably SCLs.

Z & P (p. 504) note that arbitrary gaps are more typical of affixes than clitics. There are no gaps in the set of weak subject pronouns (cf. (16)).[16]

Z & P (p. 504) note that morphological and semantic idiosyncracies are more characteristic of affixed words than

of clitic groups. There are no such idiosyncracies associated with the weak subject pronouns.

They claim (Z & P, 504) that syntactic rules can affect affixed words, but cannot affect clitic groups. The weak subject pronouns appear to fail this diagnostic, since they are available to syntactic rules of movement, as in sentential complements. In fact, it is precisely their word-like status that makes them available to movement.

By most of the criteria reviewed above, the weak subject pronouns in Somali have the characteristics of clitic pronouns. The only feature that does not fit these diagnostics is the ability to move and to remain 'independent', i.e. not cliticize. This feature is obvious in sentential complements, and I shall argue in section 4 that movement occurs as well in the case of NP focus, if the pronoun is present. That being said, I shall return to using the term 'subject clitic' (SCL) for the Somali 'weak' subject pronoun.

In this section, I have presented a preliminary analysis of Somali 'weak' subject pronouns, which have been shown to be SCLs, comparable to clitic pronouns in such better-known languages as French and Spanish. The analysis to this point has been theory-neutral. In the next section, I shall continue the analysis of SCLs, extending the discussion within a government binding framework to subject-verb agreement processes and null subjects.

Notes

1. When capitalized, 'Case' will refer to grammatical case.

2. The terms 'extensive' and 'restrictive' are due to Andrzejewski (1964).

3. The facts are actually quite complex. The paradigms in (3) apply to the matrix clause only. There are four more paradigms: extensive and restrictive paradigms for subordinate clauses that are in subject arguments of the matrix clause, and extensive and restrictive paradigms for subordinate clauses that are in non-subject arguments of the matrix clause. I reproduce from Livnat (1984, 59) the paradigms for the verb <u>keen</u> ('bring') in the present tense general (simple):

(i) Matrix clause

	extensive	restrictive
1s	keenaa	keena
2s	keentaa	keena
3sm	keenaa	keena
3sf	keentaa	keenta
1p	keennaa	keenna
2p	keentaan	keena
3p	keenaan	keena

(ii) Subordinate clause: subject

	extensive	restrictive
1s	keenaa	keenaa
2s	keentaa	keenaa
3sm	keenaa	keenaa
3sf	keentaa	keentaa
1p	keennaa	keennaa
2p	keentaan	keenaa
3p	keenaan	keenaa

(iv) Subordinate clause: non-subject

	extensive	restrictive
1s	keeno	keena
2s	keento	keena
3sm	keeno	keena

3sf	keento	keenta
1p	keenno	keenna
2p	keentaan	keena
3p	keenaan	keena

Somali verb morphology is discussed in detail in Andrzejewski (1956, 1968, and 1969).

4. There is, in fact, a range of pre-verbal particles and clitics, of which waa and the SCL form the outer layer. Hyman (1981, 176) provides an inventory of this pre-verbal material. He attributes the information to F. Antinucci:

1	2	3	4	5	6	Verb
soo	la	i	u	ma	soo	
waa	aan	ku	ka		sii	
ma	uu	is	la			
ha	ay	etc.	etc.			
etc.	etc.					

1. indicators (focused modality markers)
2. impersonal subject markers and short subject pronouns (i.e. subject clitic pronouns--APH)
3. object pronouns
4. prepositions
5. negation marker
6. deictic marker

See also Saeed (1987, 200-1)

5. El-Solami-Mewis (1987, 88) claims that ayaa may precede or follow the focused NP. This variation is not attested elsewhere, to my knowledge, and presumably reflects dialect variation.

6. Underlining in the gloss indicates the focused element. For the most part, I shall not indicate focus in the gloss, unless it is relevant to the discussion.

7. Note that the focused subject ninkii has non-nominative Case. This will be discussed below.

8. See Livnat (1984, 87-94 and 123, no. 26). Also see Saeed (1984, 42-77).

9. Somali also has direct and indirect object markers in the pre-verbal comples, which may be viewed as clitics or inflection, a question that will not be addressed here. See note 4 above for the location of these elements.

i. El-Somali-Mewis (1987, 165)

	Direct Object	Indirect Object
1s	i	ii
2s	ku	kuu
3sm	--	u
3sf	--	u
1p(inc)	ina	inoo
1p(exc)	na	noo
2p	idin	idiin
3p	--	u

The direct object paradigm is defective, lacking third person.

10. However, Roberge (1986c, 91f.) argues that the term "clitic pronoun" represents a basic linguistic category.

11. See also note 4 above.

12. Andrzejewski (1961) reports that full pronouns may also take the remote determiner, e.g. anigíi, adigíi, isagíi, iyadíi, etc. However, the proximate determiner seems to be much more frequent.

13. See notes 4 and 8 above.

14. M & S employ the terms 'stressed' and 'unstressed' for full and clitic pronouns.

15. This was brought to my attention by M. Baker.

16. The gaps in the direct object paradigm, therefore, make these latter relatively affix-like, as noted above.

3. Subject Clitics and Agreement

In section 2, I presented evidence that so-called 'weak' subject pronouns in Somali are actually SCLs. In this section, I shall explore the relationship of the SCL to the subject position and agreement processes.

3.1 Subjects and the SCL

Let us begin by considering the contrasts in (1) to (4), with IP and CP indicated. The FP+SCL complex is underlined.

(1)

a. [$_{IP}$ hooyaday <u>waa</u> karisay soorta]
mother-my F cooked cornmeal-the
'My mother cooked the cornmeal'

b. [$_{IP}$ hooyaday <u>way</u> karisay soorta]
mother-my F-she cooked cornmeal-the
(same meaning)

(2)

a. *[$_{IP}$ <u>waa</u> karisay hooyaday soorta]
F cooked mother-my cornmeal-the
(same meaning)

b. [$_{IP}$ <u>way</u> karisay hooyaday soorta]
F-she cooked mother-my food-the
(same meaning)

(3)

a. [$_{CP}$ shabeelka <u>baa</u> [$_{IP}$ nimanku dileen]]
leopard-the F men-the killed
'The men killed the leopard'

b. [$_{CP}$ shabeelka <u>bay</u> [$_{IP}$ nimanku dileen]]
leopard-the F-they men-the killed
(same meaning)

(4)

a. *nimanku [$_{CP}$ shabeelka <u>baa</u> [$_{IP}$ dileen]]
men-the leopard-the F killed
(same meaning)

b. nimanku [$_{CP}$ shabeelka <u>bay</u> [$_{IP}$ dileen]]
men-the leopard-the F-they killed
(same meaning)

<u>way</u> represents the verb FP <u>waa</u> + the 3sf SCL <u>ay</u>; <u>bay</u> represents the NP FP <u>baa</u> + 3p SCL <u>ay</u>. In (1) and (3), the SCL is optional; in (2) and (4) it is obligatory, or the sentence is ungrammatical. In (3) and (4), <u>shabeelka baa</u> or <u>shabeelka bay</u> is assumed to have moved to 'COMP',[1] in a manner that will be discussed in section 4.

The contrasts in (1) to (4) provide strong evidence that Somali is subject-initial, a matter that is in fact uncontroversial. In (1) and (3), the SCL is optional, evidently because the subject is in its D-structure position. In (2) and (4), the SCL is obligatory. Therefore, the subject has been moved to post-verbal position in (2) and to a position in front of the focused object in (4), creating structures like (5) and (6):

(5) (=(2))

a. *[$_{CP}$ e_i waa karisay hooyaday$_i$ soorta]
F cooked mother-my cornmeal-the
'My mother cooked the cornmeal'

[$_{IP}$ e_i way karisay hooyaday$_i$ soorta]
F-she cooked mother-my food-the
(same meaning)

(6) (=(4))

a. *nimanku$_i$ [$_{CP}$ shabeelka baa [$_{IP}$ e_i dileen]
men-the leopard-the F killed
'The men killed the leopard'

b. nimanku$_i$ [$_{CP}$ shabeelka bay [$_{IP}$ e_i dileen]]
men-the leopard-the F-they killed
(same meaning)

It appears therefore that the SCL 'licenses' the empty category in subject position, a topic I shall return to in 3.3. Before proceeding, however, we need to determine the precise location of the SCL. So far, we have seen that Somali permits clitic doubling of lexical NP subjects, as in (1)b and (3)b. Therefore, the SCL cannot itself be in subject position. As well, the SCL plays some role in 'licensing' the empty category left when the subject is moved, as in (2)b and (4)b. What is the position from which this might occur? The obvious candidate is I, which m-commands the subject position. For concreteness, I shall assume that the SCL is generated under AGR, from which position it governs the subject. This assumption is consistent with much recent work on SCLs.[2] Further evidence for generating the SCL under AGR will appear when we examine the interaction of the SCL with subject-verb agreement and nominative Case assignment.

3.2 Strong and Weak Agreement

It was remarked in 2.1, that Somali has a contrast between strong and weak subject-verb agreement. Weak agreement occurs with short subject extraction from matrix or relative clause;[3] strong agreement occurs elsewhere. In order to see how this pattern works, let us consider short subject extraction by WH-movement, focus, and relativization, as in (7) to (9). In the glosses, 'e' and 'r' identify extensive and restrictive verb paradigms, indicating relatively strong and weak subject-verb agreement respectively.

(7) Wh-extraction of subject

a. kumaa akhriyey buugga
who-F read(r) book-the
'Who read the book?'

b. Livnat (1984, 109)

yaa qoray warqadda
who wrote(r) letter-the
'Who wrote the letter?'

c. Saeed (1987, 223)

ninkee ayaa yimi
man-which F came(r)
'Which man came?'

(8) Subject NP focus

a. nimanka baa dilay shabeelka
men-the(-nom.) F killed(r) leopard-the
'The men killed the leopard'

b. Livnat (1984, 61)

Cali baa keena bariiska
Ali(-nom.) F bring(r) rice-the
'Ali brings the rice'

c. Saeed (1987, 217)

nimankíi ayaa keenay
men-the(-nom.) F brought(r) (it)
'The men brought it'

(9) Relativization of subject

a. [inantíi buugga akhrisay] waxay
girl-the(-nom.) book-the read(r) what-she

aragtay askari
saw soldier
'The girl who read the book saw a soldier'

b. Livnat (1984, 55)

[ninkíi naagta arkay] wuu qoslay
man-the(-nom.) woman-the saw(r) F-he laughed
'The man who saw the woman laughed'

c. Saeed (1987, 231)

nimankíi	keenay
men-the(-nom.)	brought(r) (it)

'the men who brought it'

Short subject extraction is characterized by three features: absence of SCL, non-nominative Case on the subject, and restrictive verb paradigm, reflecting relatively weak subject-verb agreement. Compare (10)a to (10)b-e, in which the focused subject <u>nimanka</u>/<u>nimanku</u> and the FP <u>baa</u>/<u>bay</u> are assumed to be extracted to COMP. The ungrammatical elements are underlined:

(10)

a.	nimanka	baa	dilay	shabeelka
	men-the(-nom.)	F	killed(r)	leopard-the

'The men killed the leopard'

b.	*nimanka	baa	<u>dileen</u>	shabeelka
	men-the(-nom.)	F	killed(e)	leopard-the
c.	*nimanka	b<u>ay</u>	dilay	shabeelka
	men-the(-nom.)	F-they	killed(r)	leopard-the
d.	*nimank<u>u</u>	baa	dilay	shabeelka
	men-the(+nom.)	F	killed(r)	leopard-the
e.	*nimank<u>u</u>	b<u>ay</u>	<u>dileen</u>	shabeelka
	men-the(+nom.)	F-they	killed(e)	leopard-the

(10)a-e have the same purported meaning. (10)b is ungrammatical because the verb is from the extensive paradigm; (10)c has a SCL; (10)d has nominative Case on the focused subject; and (10)e has all three.

Let us now review WH-extraction, NP focus, and relativization of <u>non</u>-subject NP, as in (11) to (13). In these examples, the subject is in its D-structure position. The SCL, if present, will coalesce with the FP or WH-word.

(11) Non-subject WH-extraction

a. muxuu Cali keeni doonaa
what-he Ali(+nom.) bring will(e)
'What will Ali bring?'

b. Livnat (1984, 114)

yay nimanku lacag siiyan
who-they men-the(+nom.) money give(e)
'Who do the men give money to?'

c. Livnat (1984, 111)

sannadmuu Cali dhashay
year-Q-F-he Ali born
'In what year was Ali born?'

(12) Non-subject NP focus

a. shabeelka baa/bay nimanku dileen
leopard-the F F-they men-the(+nom.) killed(e)
'The men killed the leopard'

b. Livnat (1984, 25)

muuskii baa/buu Cali cunay
banana-the F F-he Ali(+nom.) ate(e)
'Ali ate the banana'

c. Saeed (1984, 39)

lacagtíi baa naagtu keentay
money-the F woman-the(+nom.) brought(e)
'The woman brought the money'

Note that (12)c would be equally acceptable with <u>bay</u> (=<u>baa</u> + 3sf SCL <u>ay</u>), although, as it happens, such an example is not included by Saeed.

(13) Relativization of a non-subject

a. [inantíi (uu) askarigu arkay] akhrisay
girl-the (he) soldier-the(+nom.) saw(e)] read

buugga
book-the
'The girl who the soldier saw read the book'

b. Livnat (1984, 73)

[dameerkíi (uu) ninku xaday]
donkey-the (he) man-the(+nom.) stole(e)

wuu cararay
F-he ran away
'The donkey which the man stole ran away'

c. Livnat (1984, 73)

[cuntada (uu) Cali karinayaa]
food-the he Ali(+nom.) cooking(e)

way fiicantahay
F-she good
'The food that Ali is cooking is good'

When a non-subject is extracted, the extensive paradigm is selected, and the subject is assigned nominative Case. The SCL is optional if a full lexical NP is present in subject position, except in cases of WH-questions, where it appears to be obligatory, as in (11).

These observations may be summarized as in (14):

(14)

	subject extraction	elsewhere
extensive paradigm	-	+
SCL (optional)	-	+
nominative	-	+

I shall interprete this pattern to mean that Somali has two kinds of subject-verb agreement. The first, I shall call

'strong' agreement, abbreviated AGR_S. AGR_S selects verb morphology from the extensive paradigm, optionally generates a SCL, and assigns nominative Case to the subject. The second kind may be termed 'weak' agreement, or AGR_W. It selects verb morphology from the restrictive paradigm, does not generate a SCL, and does not assign nominative Case.[4] Either AGR_S or AGR_W may be generat. d under I, head of IP.[5]

The distinction between AGR_S and AGR_W accounts for when the SCL is permitted. It does not, however, account for the positive requirement that in certain contexts the SCL must occur.

3.3 SCL and pro

Consider a simple sentence lacking full lexical subject, as in (15)a, with the S-structure (15)b:[6]

(15)

a. wuu arkay shabeel
F-he saw-3sm leopard
'He saw a leopard'

b. $[_{IP}$ e $[_{I'}$ wuu $[_{VP}$ arkay shabeel]]]

(15) contrasts with ungrammatical (16), where the SCL is absent:

(16)

a. *waa arkay shabeel
F saw-3sm leopard
'He saw the leopard' (purported meaning)

b. *$[_{IP}$ e $[_{I'}$ waa $[_{VP}$ arkay shabeel]]]

What is the nature of the empty category in (15)? It cannot be a variable or NP-trace, since the only candidate for an antecedent is the SCL, and I assume, following Sportiche

(1983, see p. 207) that SCL and AGR are outside the A/A' system. Also, it cannot be PRO, because the position is governed. Therefore, it must be pro, the governed null pronominal.

In Chomsky (1982, 85-86), it is claimed that pro must be 'identified' by 'rich' agreement. More recently, Rizzi (1986) distinguished the formal licensing of pro and the mechanism by which its content is recovered. Rizzi proposes that pro is licensed by a governing head X°_{γ}, where the class of governing heads may vary from language to language. Pro is assigned content by inheriting the grammatical specifications of the governing head with which it is coindexed. In the case of strong agreement, these are the person-number features of AGR_S.[7] Now recall, from 2.1, that the SCL and extensive verb paradigm conjointly serve to uniquely identify every person-number, as in (17):

(17)	1s	aan	arkay
	2s	aad	aragtay
	3sm	uu	arkay
	3sf	ay	aragtay
	1p(inc)	aynu	aragnay
	1p(exc)	aanu	aragnay
	2p	aydu	aragteen
	3p	ay	arkeen

We may say, therefore, that pro is licensed in (15) because it is governed by AGR_S, which licenses pro in Somali. It is given content by the presence of the SCL and verb morphology from the extensive paradigm, which together identify the subject of (15) as 3sm.

Now consider the cases where a lexical NP subject has been moved to some other position (not extracted to COMP) within the clause, as in (2), repeated below as (18) with structure (19), and in (4), repeated below as (20) with the structure (21):

(18)

a. *waa karisay hooyaday soorta
F cooked mother-my food-the
'My mother cooked the cornmeal' (purported meaning)

b. way karisay hooyaday soorta
F-she cooked mother-my food-the
(same meaning)

(19)

a. *$[_{IP}$ e_i $[_{I'}$ waa $[_{VP}$ karisay hooyaday$_i$ soorta]]]

b. $[_{IP}$ e_i $[_{I'}$ way $[_{VP}$ karisay hooyaday$_i$ soorta]]]

(20)

a. *nimanku shabeelka baa dileen
men-the leopard-the F killed
'The men killed the leopard' (purported meaning)

b. nimanku shabeelka bay dileen
men-the leopard-the F-they killed
(same meaning)

(21)

a. *nimanku$_i$ $[_{CP}$ shabeelka$_j$ baa $[_{IP}$ e_i dileen e_j]]

b. nimanku$_i$ $[_{CP}$ shabeelka$_j$ bay $[_{IP}$ e_i dileen e_j]]

In (18)/(19), the subject has been moved to post-verbal position, leaving an empty category. In (20)/(21), the subject has been moved to the left of the focused object. I shall refer to this kind of movement as 'clause-bound NP movement' or 'clause-bound movement', to distinguish it from more familiar types of NP movement.

Evidence that this type of movement is clause-bound may be derived from sentential complements. In (22)b, the subject of the sentential complement, <u>nimanku</u>, has been moved to the left of the complementizer <u>in</u>, but it cannot move to the left of the matrix verb.

(22) Livnat (1984, 10)

a. Cali wuxuu moodayey [inay <u>nimanku</u> tageen][8]
Ali F-he thought that-they men-the left
'Ali thought that the men left'

b. Cali wuxuu moodayey <u>nimanku</u> [inay tageen]
Ali F-he thought men-the that-they left
(same meaning)

Examples (1) and (10) of section 1, repeated below as (23), demonstrate that clause-bound movement is also available to objects. (23)a-e have the same meaning; the object is underlined.

(23)

a. ninkii wuu arkay <u>naagtii</u>
man-the F-he saw woman-the
'The man saw the woman'

b. ninkii <u>naagtii</u> wuu arkay

c. <u>naagtii</u> ninkii wuu arkay

d. <u>naagtii</u> wuu arkay ninkii

e. wuu arkay ninkii <u>naagtii</u>

That an adjunct or adverbial expression may undergo clause-bound movement is demonstrated in (24). Note that the preposition <u>ku</u> occurs in the pre-verbal complex,[9] while the object occurs elsewhere in the clause. (24)a-c have the same meaning; the object is underlined.

(24)

a. Axmed waa ku cabay sigaar <u>beerta</u>
Ahmed F in smoked cigarette garden-the
'Ahmed smoked a cigarette in the garden'

b. Axmed <u>beerta</u> waa ku cabay sigaar
Ahmed garden-the F in smoked cigarette

c. Axmed waa ku cabay <u>beerta</u> sigaar
Ahmed F in smoked garden-the cigarette

Returning now to (18) to (21), the contrast in (18)/(19) indicates that the SCL must be present when the subject is clause-bound moved to post-verbal position. Likewise, in (20)/(21) the SCL must be present when the subject has been moved to the left of the focused object, although in this case the SCL is associated with the FP baa in COMP.[10]

What is the status, then, of the null subject indicated in (19) and (21)? The minimal assumption would seem to be that it is pro, as in (15). First, however, consider the alternatives. It cannot be PRO, because the empty category is the subject of a finite clause and therefore is governed. It cannot be NP-trace, because the antecedent is not in an A-position. One can make a case that it is a variable, on the assumption that the antecedent is in an A'-position. However, this approach has some undesirable consequences. The core instance of variable is created by extraction to COMP. However, I have already indicated that subject extraction is only possible from the domain of AGR_W.[11] The clause-bound type of subject NP movement being considered here occurs only within the domain of AGR_S, and has the additional requirement that the SCL be present. Therefore, if the empty subject position in (18) to (21) is an A'-bound variable, we must distinguish two quite distinct kinds of variable. The first, the core case, is created by movement to COMP from the domain of AGR_W, with the possibility of extraction from sentential complement to matrix clause via COMP-to-COMP movement. The second type of variable would be created by local, clause-bound movement to nearly any position except COMP, from the domain of AGR_S (not AGR_W), with the additional requirement that the SCL be present. Positing that the empty subject position in (18) to (21) is a variable therefore entails that Somali has two kinds of variable which are in virtual complementary distribution, indicating that the approach is untenable.

What remains, then, is the alternative that the moved NP leaves a pro, which is licensed by virtue of the position it occupies and is given content when SCL is present. At this point, a careful distinction is required. Brody (1984) has argued against changing the categorial determination of an empty category in the course of a derivation. If we accept his argument, then in the present case we must say that clause-bound NP movement in (18) to (21) leaves a pro; we must avoid saying that it leaves a trace or variable which is reinterpreted as pro at a later step in the derivation.

Before proceeding, let us briefly consider the level at which clause-bound NP movement occurs. Although the evidence is less than conclusive, it is likely that this movement occurs at PF. In the first place, clause-bound movement is strictly local, so there is an absence of the kind of cyclic movement typical of major projections at S-structure, e.g. WH-movement. Secondly, clause-bound movement is very free with respect to landing site, and in particular, there is no evidence that clause-bound movement feeds binding theory.[12]

If this type of movement occurs at PF, it entails that the requirement that pro be licensed and be given content must occur at that level, as well as perhaps at other levels of the grammar. This seems unproblematic. The requirement that pro be licensed and identified should be construed as a well-formedness condition which applies wherever relevant.

We have now examined null subjects and pro created by clause-bound NP movement. The third context we shall now consider in which SCL is obligatory is the case of full lexical pronouns.

In 2.2, I observed, following Kayne (1975) and Jaeggli (1982), that in languages having both clitic and full lexical pronouns, the latter tend to be [+animate]. I also observed that such full lexical pronouns are generally clitic-doubled. We now need to account for why the SCL is obligatory in Somali when the subject position is occupied by a full lexical

pronoun.[13] For convenience, I repeat the Somali data from 2.2:

(26)

a. gurigii wuu gubtay
house-the F-he burned
'The house burned'

b. wuu gubtay
F-he burned
'It (the house) burned'

c. *isagu wuu gubtay
he F-he burned
'It (the house) burned'

(27)

a. faraskii wuu turanturooday
horse-the F-he tripped
'The horse tripped'

b. wuu turanturooday
F-he tripped
'He (the horse) tripped'

c. isagu wuu turanturooday
he F-he tripped
'He (the horse) tripped'

(28)

a. isagu wuu arkay
he F-he saw
'He saw (him/it/her/them)'

b. *isagu waa arkay
he F saw
(same meaning)

(29) Livnat (1984, 24)

a. ninkii baan anigu arkay
man-the F-I I saw
'I saw the man'

b. *ninkii baa anigu arkay
man-the F I saw
(same meaning)

(26) and (27) illustrate that full lexical pronouns in Somali are [+animate]. Some speakers of Somali further restrict full pronouns to [+human]. (28) and (29) illustrate obligatory clitic-doubling of full subject pronouns in Somali.

In order to extend the present approach to this data, it is necessary to assume the Avoid Pronoun Principle of Chomsky (1981, 65), also discussed in Jaeggli (1982, 42 and 93). The Principle might be expressed, within current versions of the theory, as (30):

(30) Avoid pronoun if pro is possible.

In English, pro is not possible, so a full lexical pronoun is obligatory. In Somali, pro is possible in subject position, so (30) applies. How is it, therefore, that a full subject pronoun can appear, why must it be clitic-doubled, and why is it [+animate] and relatively emphatic?[14]

Let us assume that (30) applies at D-structure, so pro is inserted at D-structure in Somali, at which level it must also be licensed and given content. I propose that pro may optionally be specified [+/-emphatic] at D-structure. If [-emphatic] (or simply unspecified), pro behaves in the normal fashion, hence the typical case where there is a SCL and no full subject pronoun. If [+emphatic] at D-structure, pro is [+emphatic] at all levels of the derivation. It is reasonable to suppose that a category that is [+emphatic] cannot be phonetically null at PF. Therefore, at PF [+emphatic] pro will be given phonetic content as a full pronoun. The feature [+emphatic] might be associated with the features [+animate]/[+human] at some pragmatic level of the grammar, depending in part on the significance one attaches to slight linguistic variation.[15]

Given these minimal assumptions, it follows naturally that a full lexical pronoun in Somali will be clitic-doubled

and [+animate], as appears to be the case in analogous structures in Romance languages as well.

We have now dealt with three contexts in which the SCL is obligatory in Somali: when the subject position is null (i.e. pro), when the subject undergoes clause-bound movement leaving pro, and when a base-generated pro marked [+emphatic] at D-structure is spelled out at PF as a full lexical pronoun. In each case we have found the subject position to be occupied by pro at some level of the derivation. This is virtually a definition of a null subject language. It is now appropriate to compare the Somali facts with generalizations that have been made with respect to null subject languages. We shall see that Somali is both like and unlike the more familiar examples.

3.4 Somali as a Null Subject Language

The literature on null subject languages is extensive and complex, and it is beyond the scope of the present study to provide a complete history of the subject. However, it is generally agreed that the topic was introduced to the transformational-generative literature in Perlmutter (1971), where it was observed that null subject languages do not exhibit what have come to be known as 'that-t' effects. The early work was limited to 'classic' null subject languages like Standard Italian, where verb inflection encodes person-number features, permitting the absence of a lexical subject. Languages with SCLs were not counted as members of this class, probably due to the influence of Kayne (1975). Kayne proposed that SCLs in Standard French were generated in subject position and moved transformationally to VP, it being his intent to capture the complementary distribution of SCLs and full lexical subjects in Standard French. Later, Jaeggli (1982) observed that some dialects of Spanish permit clitic-doubling of objects, and he proposed an analysis generating

object clitics on the verb, to account for these facts. More recently, Roberge (1986a, 1986b, 1986c) has observed that some dialects of non-standard French permit clitic-doubling of subjects. He therefore proposes to generate SCLs under AGR. When a full lexical subject is absent, the subject position is pro, which is given content by the SCL. Somali must be understood as a null subject language in the sense proposed by Roberge, i.e. with pro subject and either 'rich' verb morphology or a SCL.

Chomsky (1981) proposes what are perhaps the most comprehensive generalizations concerning null subject languages (PRO-drop languages, in the terminology then current). These, he claims, share the properties listed in (31):

(31) Chomsky (1981, 240)

a. missing subject
b. free inversion in simple sentences
c. 'long wh-movement' of subject
d. empty resumptive pronouns in embedded clause
e. apparent violations of *[that-t] filter

Corresponding to these respective properties are examples drawn from Italian, as in (32):

(32) Chomsky (1981, 240)

a. ho trovato il libro
'(I) found the book'

b. ha mangiato Giovanni
'Giovanni ate'

c. l'uomo [che mi domando [chi abbia visto]]
(with the interpretation: 'the man x such that I wonder who x saw')

d. ecco la ragazza [che mi domando [chi crede [che possa VP]]]
'This is the girl who I wonder who thinks that she may VP'

e. chi credi [che partirà]
'who do you thing (that) will l ave'

I shall now review how Somali coincides with these purported properties of null subject languages.

First, Somali clearly permits missing subjects, i.e. (31)a, as we have seen in (15) above.

Second, Somali does <u>not</u> permit free inversion, (31)b. I shall return to this directly.

Third, Somali permits long WH-movement, (31)c. The following example is a case of relativization, but the same principle is at work:

(33) Livnat (1984, 37)

naagta uu [Cali qabo [inay ninka buugga siisay]]
woman-the he Ali thinks that-she man-the book-the gave

waa Amina
F Amina

'The woman who Ali thinks (that) gave the book to the man is Amina'

Fourth, Somali has empty resumptive pronouns in embedded clause, (31)d. This is illustrated in (33), also (34):

(34) (=(6)a of 2.1)

Cali$_i$ baan [e mooday [$_{CP}$ inuu e$_i$ Berbera ka yimid]]
Ali F-I thought that-he Berbera from came
'I thought that <u>Ali</u> came from Berbera'

<u>Cali</u>, subject of the sentential complement, has been focused, leaving an empty category, which I shall argue in section 5 is a resumptive null pronominal pro. Discussion of this point must be deferred until then.[16]

Finally, Somali has apparent violation of that-t effects, (31)e. This is amply illustrated in (32) and (34). In both cases, an embedded subject has been extracted over the complementizer <u>in</u>.

Returning to the question of free inversion, this property is crucial to the way null subject languages generally have been understood. Since Rizzi (1982, c.4), it has been widely accepted that subject extraction in null subject languages like Italian is from post-verbal position, following free inversion, which permits apparent violation of the that-t filter (or ECP, in current versions of the theory). Therefore, finding a null subject language with apparent violation of the that-t filter, which lacks free inversion, is of some interest. There is prima facie evidence, however, that Somali has free inversion in such examples as (35), derived from (18) and (19) above:

(35)

a. way karisay hooyaday soorta
F-she cooked mother-my cornmeal-the
'My mother cooked the cornmeal'

b. $[_{IP}\ e_i$ way $[_{VP}$ karisay hooyaday$_i$ soorta]]

In (35), the subject has been moved to post-verbal position by a process I have called clause-bound NP movement. This process leaves an empty subject, which I have argued is pro. How is this to be distinguished from 'true' free inversion? In order answer this question, we must review the crucial properties of free inversion, using examples from Standard Italian and the Northern Italian dialects of Fiorentino and Trentino.

Italian permits a missing subject, as in (36)a, where verb inflection encodes the person-number features of the subject.

(36) Standard Italian

a. ho trovato il libro
'(I) found the book'

b. ha mangiato Giovanni
has eaten Giovanni
'Giovanni ate'

In free inversion, as in (36)b, the subject appears in immediate post-verbal position and the verb agrees with the inverted subject. There have been a variety of derivations proposed for this construction. One derivation is to generate the subject in [NP,IP] and move it to post-verbal position at S-structure. Alternatively, the post-verbal subject may be generated _in situ_. In either case, it is assumed that the subject position is occupied by a pleonastic null pronominal (PRO in Chomsky (1981), pro in more recent versions). This assumption is required, in part, because the null subject would otherwise bind the post-verbal subject, creating a Principle C binding theory violation. The analysis assumes that free inversion occurs at S-structure and feeds binding theory.

The Northern Italian dialects differ from Standard Italian in having obligatory SCLs. When the subject is in [NP,IP], the SCL and verb agree with the subject. In free inversion, Fiorentino displays a neuter clitic and the verb agrees with the pleonastic null subject, not with the post-verbal subject, as in (37):

(37) Fiorentino (Roberge 1986c, 122)

a. Le ragazze le vengano
the girls they come-3p
'The girls come'

b. E' viene le ragazze
it come-3s the girls
(same meaning)

In (37)a, the SCL and verb agree with the subject, _le ragazze_. In free inversion, as in (37)b, the neuter SCL _e_ appears, and the verb is 3s.

Turning to Trentino, the SCL paradigm in this dialect is defective, lacking a neuter SCL corresponding to e in Fiorentino. Hence, there is no SCL in free inversion. Otherwise, Trentino patterns with Fiorentino:

(38) Trentino (Roberge 1986c, 122)

a. 'Na putela l' ei vegnuda
a girl she has come
'A girl came'

b. E vegnu 'na putela
has come a girl
(same meaning)

We may now summarize the characteristics of free inversion in Standard Italian and the Northern Italian dialects:

(39) Properties of free inversion

a. Specific to subjects
b. Subject appears in immediate post-verbal position
c. [NP,IP] occupied by a pleonastic null pronominal, which is not coreferential with the subject
d. The subject is generated in situ, or is moved at S-structure. In either case, free inversion feeds binding theory
e. In dialects with SCLs, the SCL and verb agree with the pleonastic subject; in dialects without SCLs, the verb agrees with the post-verbal subject

Now, compare these with Somali:

(40) Clause-bound NP movement in Somali

a. Applies freely to all NPs
b. Subjects may appear in a wide variety of positions, not only post-verbal position
c. If the subject is moved, [NP,IP] is occupied by pro, which is co-referential with the moved subject.
d. Movement occurs at PF and does not feed binding theory
e. SCL is obligatory if the subject is moved from [NP,IP]. The SCL and the verb agree with the moved subject

The evidence indicates that Somali, although it is a null subject language, lacks free inversion. In this respect, Somali patterns with Portuguese. Roberge (1986c, 115-6), drawing upon Perlmutter (1976), provides the following evidence:[17]

(41) Portuguese

a. Deus existe
'God exists'

b. *Existe Deus

No attempt will be made here to systematically account for these variations. It will suffice for the present to establish that Somali is a null subject language lacking free inversion, a point that will be relevant to the discussion of subject extraction in section 5.

3.5 Summary

In this section, I have argued that Somali has a contrast between strong and weak subject-verb agreement, AGR_S and AGR_W respectively. AGR_S generates an optional SCL, assigns nominative Case, and selects verb morphology from the extensive paradigm. AGR_W does not generate a SCL, does not assign nominative Case, and selects verb morphology from the restrictive paradigm. AGR_W occurs with short subject extraction from matrix or relative clause; AGR_S occurs elsewhere.

Three contexts have been examined where the SCL is obligatory: where there is no overt subject, where the subject has been moved (other than being extracted), and where the subject is a full lexical pronoun. It is argued that in all these cases, [NP,IP] is occupied by pro at some level of the derivation. Hence, the SCL contributes to giving pro content, a recognized requirement for pro.

Having established that Somali is a null subject language, I presented evidence that the language lacks free inversion, a property that has frequently been associated with null subjects.

Notes

1. The term 'COMP' is out of place within the Barriers framework. I am using it in a generic sense for the complementizer position until section 4, where a specific analysis will be proposed.

2. See Jaeggli (1984), Roberge (1986a, 1986b, 1986c), Everett (1987). Note, however, that Kayne (1975) argues for French SCLs in subject position, primarily to account for the complementary distribution of SCLs and lexical NP. According to Safir (1986), SCLs in Trentino are on VP.

3. Subject extraction from a sentential complement, however, is from the domain of strong agreement. I shall discuss this phenomenon at length in section 5.

4. There is some evidence for a contrast between AGR_S and AGR_W in other Afro-Asiatic languages. Owens (1985, 108-110) reports that Oromo permits an 'emphatic' subject, which I assume to be the equivalent of a focused subject in Somali. The emphatic subject has distinctive Case morphology, and in the context of an emphatic subject, verb morphology is neutralized to 3sm. See also Kenstowicz (1989) on Arabic.

5. In section 4, I shall propose that the FP is also generated under AGR.

6. FP waa + SCL uu --> wuu. The assumption that waa is generated under I will be discussed in section 4.

7. Jaeggli and Safir (1989) propose a further refinement in the theory of null subjects. Null subjects are said to be permitted 'in all and only languages with morphologically uniform inflectional paradigms.' (J and S 1989, 29), where morphological uniformity is defined:

i. Morphological Uniformity (J and S 1989, 30)

An inflectional paradigm P in a language L is morphologically uniform iff P has either only underived inflectional forms or only derived inflectional forms.

In a language with only derived inflectional forms, pro will be identified through government by 'rich' or 'strong' AGR. This is the case of Italian and Somali. In a language like Chinese, with only underived inflectional forms, AGR will

be morphologically null, but will inherit features from a c-commanding NP.

8. These sentences are of the type (14) in section 1. Note that waxaa+uu --> wuxuu.

9. See note 4, section 2.

10. The subject in (20)/(21) appears to be topicalized, although it is unlikely the construction is exactly like topicalized structures in English, particularly in light of the analysis of clause-bound NP movement offered below.

11. That is to say, from the domain of AGR_w in the case of matrix clause and relative clause. Extraction from sentential complement is from the domain of AGR_s. I shall discuss this distinction in section 5.

12. Clause-bound movement at S-structure would require an elaboration of the theory of phrase markers. The pro left by clause-bound movement is not pleonastic, since it is obligatorily identified by the person/number features of its antecedent. Therefore, pro and its antecedent could form a chain. If the lexical subject is in post-verbal position, it would be bound by the pro in [NP,IP], a Principle C binding theory violation.

Elaborating the theory of phrase markers to eliminate this problem seems a highly undesirable move.

13. Full lexical pronouns may also occur as direct objects, in which case they are clitic-doubled (assuming the term 'clitic' is appropriate) in the first and second person. As indicated (note 4, section 2), the paradigm is defective, lacking third person.

For an example, we may turn to El-Solami-Mewis (1987, 114-7), who reports on an impersonal construction with the subject la, 'one'. In the following, la ('one') + i ('me') --> lai and la ('one') + ku ('you' 2s) --> lagu. I have translated the glosses from the German.

i. Aniga waa lai dili
Aniga F one-me beats
'One is beating me' (I am being beaten)

ii. Adiga waa lagu dili
you(2s) F one-you(2s) beats
'One is beating you' (You are being beaten)

Note that la (unlike the personal SCLs) does not coalesce with waa. (The reader may notice that the verb dil may be

translated variously as 'hit, beat' as it is here, or as 'kill' as in the text. See Luling (1987).)

14. A similar range of facts occurs in Spanish with respect to object clitics, and are discussed in Jaeggli (1982, 40-45). Consider the following data from Spanish:

(i) Jaeggli (1982, 41 and 43)

a. *Vimos él
b. Lo vimos
'We saw him'
c. Lo vimos a él
(same meaning)

(i)a-c have the same purported meaning. Jaeggli's analysis is based on the Avoid Pronoun Principle and the assumption that object clitics absorb Case. (i)a is ungrammatical by the Avoid Pronoun Principle; (i)b is the standard case and is grammatical; and (i)c is permitted by whatever permits clitic-doubling of lexical NP--in Jaeggli's analysis, by the presence of the Case assigner a.

There is no evidence that the SCL in Somali absorbs (nominative) Case: the lexical subject (including full pronoun) is assigned nominative Case by AGR_S, regardless whether SCL is present. Moreover, there is, of course, no overt Case assigner comparable to Spanish a, and in particular, there is no overt Case assigner which occurs precisely when the subject is clitic-doubled. Therefore, Jaeggli's approach will not account for the facts in Somali.

15. Recall, however, (45)a-b in section 2, in which full lexical pronouns occur in contexts where SCLs are not permitted, i.e. focused subject, and subject of a direct predication clause of the form 'NP waa X'.

16. Note the presence of the SCL uu (he) in the sentential complement. Livnat (1984, 135f.), working from different assumptions, calls this a SCL (as I do), but also claims that the SCL is itself the resumptive pronoun.

Within my approach, the SCL is generated under AGR_S, in order to account for clitic-doubling of lexical subjects. In the absence of lexical subject, I take the subject position to be occupied by pro. This pro will be obligatorily clitic-doubled. The pro-SCL configuration occurs in three contexts: a) if lexical subject is entirely absent, b) if lexical-subject is clause-bound moved, or c) if lexical subject is extracted from sentential complement via WH- or focus movement. In the latter case, pro (not the SCL) is the resumptive pronoun.

17. Roberge (1986c, 119) claims that Irish is also a null subject language lacking free inversion.

4. Focus and Word Order

In 2.1, while introducing some basic facts about Somali syntax, I stated that every indicative matrix clause, and only the matrix clause, has one, and only one, focus particle (FP). Subsequently, in 3.1, there was reason to refer to focus movement of NP and NP FP to 'COMP'. This is somewhat of a misnomer, since 'COMP' is reinterpreted in the Barriers framework as head and specifier of CP. What must be done now is to describe and account for focus phenomena, especially NP focus, and how they interact with other elements, including SCL and word order.

4.1 Focus

As indicated in 2.1, the verb FP waa occurs in the outer layer of pre-verbal particles. (See section 2, note 4.) The verb FP is illustrated in (1):

(1)

ninkii	waa	arkay	naagtii
man-the	F	saw	woman-the

'The man saw the woman.'

On the other hand, the NP FP baa or ayaa occurs after the focused NP:

(2)

a.

ninkíi	baa	arkay	naagtii
man-the	F	saw	woman-the

'The man saw the woman'

b.

naagtíi	baa	ninkii	arkay
woman-the	F	man-the	saw

'The man saw the woman'

In (2)b, the focused object and FP precede the subject. Livnat (1984, 47f.) proposes that the focused NP is 'outside the clause' (i.e. outside IP, in the Barriers framework). The present analysis follows on Livnat's insight, and it has been assumed, without argumentation, that NP focus entails movement to COMP. Accordingly, (3) represents a simplified approximation of the presumed S-structure of (2)b:

(3) $[_{CP}$ naagtii$_i$ baa $[_{IP}$ ninkii arkay e_i]]

Several questions arise: first, what is the location of waa in (1)? Second, what relationship obtains between waa, baa, and ayaa? Third, what syntactic processes underly NP focus? I shall take these in order.

On the basis of sentences like (1) and (2) of 3.1, repeated below as (4)a-b, I have argued that Somali is subject initial and that the SCL is generated under AGR_S, where it m-commands the subject position.

(4)

a. hooyaday way /waa karisay soorta
mother-my F-she/F cooked cornmeal-the
'My mother cooked the cornmeal'

b. way /*waa karisay hooyaday soorta
F-she/ F cooked mother-my cornmeal-the
(same meaning)

Given the present analysis, where is the locus of waa? The minimal assumption is that waa, like the SCL, is generated under AGR.[1] There is independent evidence to support this analysis. Somali has a range of verbless sentences, like (13)a to (13)c of 2.1, repeated below as (5), in which the focused constituent follows the FP waa.

(5)

a. walaalkay waa barre
brother-my F teacher
'My brother is a teacher'

b. tareenkii waa raagay
train-the F late
'The train is late'

c. ninka guursaday walaalshay waa alwaaxqorre
man-the married sister-my F carpenter
'The man who married my sister is a carpenter'

As the FP waa is central to this construction, it is plausible that it is generated under I, from which the clause IP is projected.

Cross-linguistic evidence is available from Hebrew. Doron (1986, 313) reports that Hebrew and related languages like Arabic have a kind of present tense nominal sentence, as in (6):

(6) Doron (1986, 313)

dani hu ha-more
Dani he the teacher
'Dani is the teacher'

Doron argues that the pronoun in such constructions is a clitic pronoun 'which is the phonological realization of "unattached" agreement features that have absorbed Case' (Doron 1986, 313). Assuming this clitic is generated under AGR, Somali and Hebrew share the feature of allowing verbless sentences in which I serves to mark a direct predication relation. They differ in the realization of I, Somali using a FP and Hebrew a clitic pronoun.

Let us assume, therefore, that waa is generated under AGR. Further, let us provisionally assume that I directly precedes V and VP (a question to which I shall return directly in section 4.3). So, the analysis of Somali phrase structure to which I am presently committed is indicated in (7).

(7) $\ldots[_{IP}$ NP $[_{I'}$ I $[_{VP}$ V ...

I shall assume that clitics 'inside' the waa+SCL complex are generated on the verb, though in fact the question is immaterial for the present discussion. Given these assumptions, the order of waa+SCL relative to the verb and other pre-verbal clitics falls out directly.

Let us now consider the possibility that baa and ayaa, like waa, are generated under AGR (both AGR_S and AGR_W), together with an optional SCL, in the case of AGR_S, then move to 'COMP' at S-structure. Similarly, the focused NP moves from its base-generated position to 'COMP', also at S-structure.

AGR is a head, so movement of FP+SCL generated under AGR is subject to the Head Movement Constraint of Travis (1984, 131), which is incorporated into the Barriers framework as (8) (Barriers, 71):

(8) Head Movement Constraint

> Movement of a zero-level category b is restricted to the position of a head a that governs the maximal projection g of b, where a theta-governs or L-marks g if a $\neq$ C.

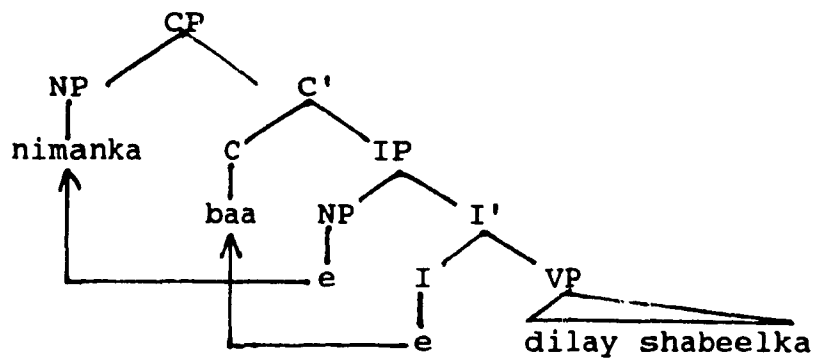
By the Head Movement Constraint, FP+SCL moves to C, head of CP, which governs IP, the maximal projection of I.

Furthermore, we may assume on general grounds that the focused NP moves to [NP,CP] at S-structure, parallel to WH-movement and relativization. The S-structures associated with subject and non-subject focus will then be as in (9) and (10).

(9) Subject focus

a. nimanka baa dilay shabeelka
men-the F killed leopard-the
'The men killed the leopard'

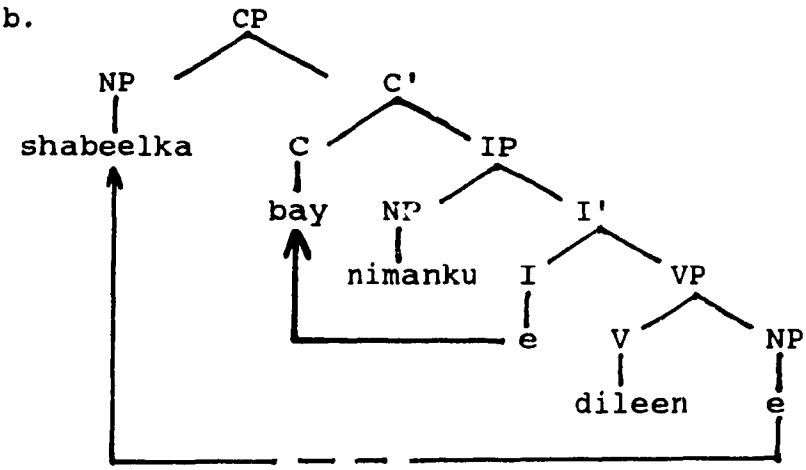
b.



(10) Non-subject focus

a. shabeelka bay nimanku dileen
leopard-the F-they men-the killed
(same meaning)

b.



In (9), nimanka m-commands and antecedent-governs its trace in [NP,IP]. Likewise, baa m-commands and antecedent-governs its trace in I.

Extraction of shabeelka in (10) is more complicated, as VP is a barrier. However, shabeelka may adjoin to VP,

overcoming barrierhood of VP, then move to [NP,CP], IP and I' not being barriers. (See Barriers, 21-22.) As to proper government of the trace, the theory is open on this question. It may be that the verb dileen theta-governs and properly governs the trace of shabeelka, or it may be that shabeelka antecedent-governs its trace. The analysis and the alternatives are virtually the same as for English.

The evidence from (10) supports a movement analysis of non-subject focus. However, what is to prevent a non-movement derivation of subject focus in (9)? That is, why are subject and FP not in situ at S-structure in this construction, along the lines of the vacuous movement hypothesis of Chomsky (1986, 48f.)? One reason for rejecting this analysis is that there are definite co-occurence restrictions on subject focus and respectively nominative Case assignment, extensive paradigm verb morphology, and SCL (all generated under AGR_S). It is natural to capture these restrictions as defining a domain that is opaque for purposes of subject extraction, as will be discussed at length in section 5. Furthermore, the vacuous movement hypothesis leaves COMP empty and therefore available for WH-movement at S-structure. However, focus constructions and WH-constructions are in strict complementary distribution in Somali, which reduces the plausibility of the subject remaining in situ in sucject focus constructions.

Movement of bay to head of CP raises some questions, as the SCL and (presumably) the trace are co-indexed with the subject, nimanku. Is the subject position governed by bay or by its trace? Adams (1987) associates the null subject parameter with directionality of government. In section 3.3 I argue that a subject pro may be licensed by SCL under AGR. Take this to indicate direction of canonical government. Given Adams' anaysis, it would follow that subject pro must always be governed from the position of AGR. In the case of

SCL in C, SCL will form a chain with a trace at the extraction site in AGR, from which pro will be governed.

Observe that focus movement in (10) is directly parallel to V2 phenomena in Germanic languages.[2] Compare (11):

(11)

a. Who did the men kill?

b. $[_{CP}$ who$_i$ $[_{C'}$ did$_j$ $[_{IP}$ the men $[_{I'}$ t_j $[_{VP}$ kill t_i]]]]]

<u>Who</u> is extracted, via adjunction to VP, to the position [NP,CP], while <u>did</u> moves to the position head of CP, a species of head movement.

NP cannot be extracted from a relative clause, presumably by the Complex Noun Phrase Constraint:

(12) Livnat (1984, 54)

a. *Axmed buu Cali arkay [naagta uu jecelyahay]
Ahmed F-he Ali saw woman-the he loves
'Ali saw the woman that <u>Ahmed</u> loves'

b. *lacag buu Cali arkay [ninka uu Axmed siiyey]
money F-he Ali saw man-the he Ahmed gave
'Ali saw the man to whom Ahmed gave <u>money</u>'

However, NP may be extracted from a sentential complement. In particular, the subject of a sentential complement may be extracted:

(13)

a. ninka$_i$ baan ramaysanahay [inuu e$_i$ tegey]
man-the F-I think that-he AGR$_S$ left(e)
'I think that the man left'

b. Livnat (1984, 43) (= (6)a of section 2)

Cali$_i$ baan mooday [inuu e$_i$ Berbera ka yimid]
Ali F-I thought that-he Berbera AGR$_S$ from came(e)
'I thought that Ali came from Berbera'

c. Livnat (1984, 16)

bariis$_i$ baan mooday [inuu ninkaani iibiyo e$_i$]
rice F-I thought that-he man-this AGR_S sells(e)
'I thought that this man sells rice'

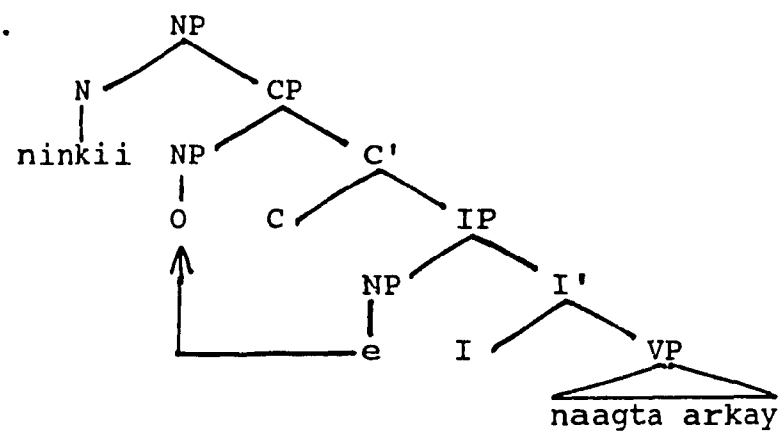
Observe that the subject is extracted from the domain of AGR_S and the presence of SCL is obligatory. These facts will be accounted for in section 5.

This analysis of focus extends naturally to relative clause formation. Consider (14), from Livnat (1984, 55), with my proposed analysis of the subject NP, and (15) from Livnat (1984, 56).

(14) Subject relativized

a. [ninkii (*uu) naagta arkay] wuu qoslay
man-the (*he) woman-the saw(r) F-he laughed
'The man who saw the woman laughed'

b.

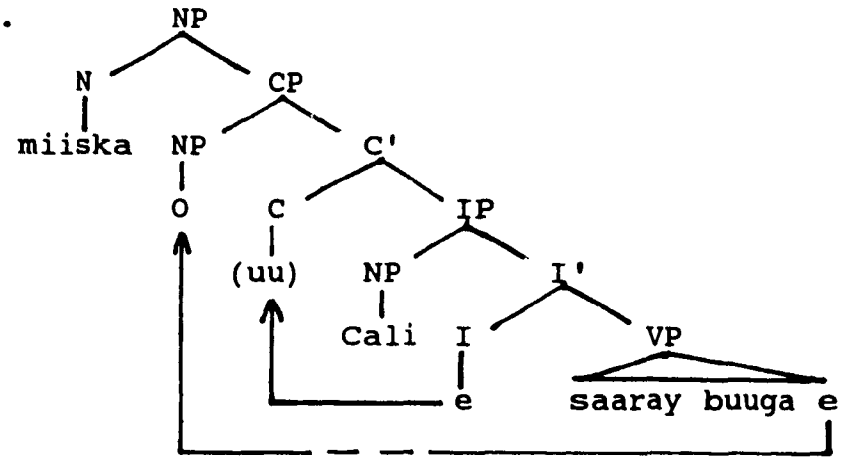


I assume that a base-generated empty operator is extracted from subject position to [NP,CP]. AGR_W selects from the restrictive verb paradigm and does not generate a SCL.

(15) Non-subject relativized

a. [miiska (uu) Cali saaray buugga] wuu weynyahay
table-the (he) Ali put(e) book-the F-he big
'The table Ali put the book on is big'

b.



Again, a base-generated empty operator is extracted, this time from an object position. AGR_S selects from the extensive paradigm, assigns nominative Case, and generates an optional SCL, which moves to head of CP. Hence, the SCL follows the head of the relative clause and precedes all other constituents in the clause (except the empty operator, which is not visible), although word order within the clause is otherwise relatively free. The SCL is optional if there is a lexical subject in its D-structure position; it is obligatory if the subject is clause-bound moved (leaving pro), if it is a lexical pronoun, or if it is base-generated pro. The same mechanisms are at work here and in the focus construction, except for the empty operator and the absence of FP.

The analysis presented so far accounts for a number of facts. First, generating waa+SCL under AGR accounts for these particles and clitics appearing in the outer layer of pre-verbal clitics (see note 4, section 2).[3] Second, generating baa/ayaa under AGR accounts for the fact that they are in complementary distribution with each other and with waa.

Third, the separate movement of baa/ayaa+SCL to head of CP and movement of focused NP to [NP,CP] accounts for these constituents being sentence-initial, for the relative order they have to each other, and for the fact that the focused NP may originate from a sentential complement. Fourth, the extension of the account to relativization accounts for the SCL phenomena in the relative clause, and in particular for the relative order of SCL to other constituents.

We can now see that the major difference between Somali and English with respect to focus is that focus movement is overt in Somali and occurs at S-structure. English exhibits focus through contrastive stress, as in (16).

(16) John likes MARY (i.e. not Joan)

In the standard theory, the semantic interpretation of English focal stress is parallel to that of quantifiers, the focused expression being raised at LF to sentence-initial position, where it acts as an operator binding a variable, as in (17) (Chomsky (1976), May (1985), Rochemont (1986), Culicover and Rochemont (1983)).

(17) (for x = Mary) $[_{IP}$ John likes x]

4.2 aa Insertion

In 2.2, I argued that 'weak' pronouns in Somali are SCLs, and in 3.2, that they are generated under AGR_S in I. I have just argued that the FPs waa, baa, and ayaa are also generated under AGR. When the FP and SCL co-occur in the same clause, they coalesce, as we have already seen. The paradigm for the various combinations is in (18):

(18)

	waa	baa	ayaa
1s	w-aan	b-aan	ay-aan
2s	w-aad	b-aad	ay-aad
3sm	w-uu	b-uu	ay-uu
3sf	w-ay	b-ay	ay-ay
1p(inc)	w-aynu	b-aynu	ay-aynu
1p(exc)	w-aanu	b-aanu	ay-aanu
2p	w-aydu	b-aydu	ay-aydu
3p	w-ay	b-ay	ay-ay

It appears that some generalization is being missed, as the FPs uniformly end in aa, and they undergo a common morphophonemic process when they coalesce, as in (18). Therefore, I want to investigate the possiblity that the underlying representations are actually w-, b-, and ay-.[4] When they oc-occur with SCLs, that is, when the FP and SCL are jointly generated under AGR_S, the representations will be as in (19):

(19)

```
w- + uu -->  wuu
FP   he

b- + uu -->  buu
FP   he

ay- + uu --> ayuu
FP    he
```

As well, some means must be available to allow w-, b- , and ay- to have well-formed surface structures in the absence of a SCL. Therefore, let us suppose there is a rule of 'aa insertion', which we might express informally as in (20).

(20) aa insertion

AGR --> $[_{AGR}$ aa]

The rule is optional and may apply under AGR_S and AGR_W. We may think of aa insertion as something like 'do support' in English. Evidently, like do support, aa insertion is a root

phenomenon, occuring only in the matrix clause, as aa does not appear in subordinate clauses, even if SCL is absent.[5]

4.3 Word Order

So far, I have addressed a number of questions regarding Somali syntax, all the while skirting the issue of word order. This may seem rather peculiar. However, given the relatively free word order in the language, the strategy I have chosen is to ascertain what can be decided with some degree of confidence, and only then address the more elusive questions. Two of the most elusive, in fact, are the location of I relative to VP and the internal structure of VP.

Consider what has been decided to this point: Somali is subject initial, FP+SCL are generated under AGR, extraction is leftward. Further, I have assumed, without explicitly addressing the question, that I and AGR immediately precede V and VP. So the phrase structure of the language, simplified, is (21).

(21) $[_{CP}\ [_{IP}\ NP\ [_{I'}\ I\ [_{VP}\ V\ NP\ ]\ ]\ ]$

The relative order of I and VP and the internal structure of VP have only been assumed, however the question does not bear directly upon subject extraction, since I governs the subject position in any case. Nevertheless, the relative order of I and VP, and the internal structure of VP are basic questions, and it is worth a digression to examine the issues and make some proposals, even if they must be in part only provisional.

It has been implicit in the argument to this point that generating w- and SCL under I, and assuming the structure (21), will account for the correct order of verb FP and SCL relative to the verb and other pre-verbal clitics, as in (1) and (4) above, and (22) and (23) below:

(22) Saeed (1987, 200)

wuu ku soo noqday
F-he to back came
'He came back to it'

(23) Saeed (1984, 170)

baabuurkii wuu i dhaafay
truck-the F-he me passed
'The truck passed me'

This approach has the merit of simplicity, but it does not provide a definitive argument for the word order claimed above, since clitic movement is not necessarily subject to strict adjacency requirements.[6]

A stronger argument for the location of I may be derived from the analysis of SCL in sentential complement, as in (24):

(24)

a. Axmed baa u malaynaya [in<u>uu</u> berito
Ahmed F thinks that-he tomorrow

booqanayo barraha]
visit teacher-the
'Ahmed thinks that he will visit the teacher tomorrow'

b. Axmed baa u malaynaya [in berito <u>uu</u> barraha booqanayo]

In sentences like (24), the SCL in the sentential complement may occur anywhere between the verb and the complementizer <u>in</u>. In particular, NP may occur between the SCL and the verb, and between SCL and <u>in</u>.

The evidence lends itself to the following analysis: Suppose that <u>w-</u> obligatorily cliticizes onto the verb at PF, carrying with it the SCL, if present. The correct relative order of <u>w-</u>, SCL other pre-verbal clitics, and the verb will fall out directly. Let us assume further that the SCL, of itself, does not cliticize onto the verb. Hence, in the sentential complement (where <u>w-</u> is absent) it may remain <u>in</u>

situ under AGR. The relatively free word order of Somali, effected by clause-bound NP movement, accounts for the possibility that one or more NPs may occur between the SCL and the verb.[7]

This analysis supports the claim that Somali is SIVP.[8] However, it is consistent with the language being either SIOV or SIVO. I have been assuming implicitly that Somali is SIVO. The evidence is hardly decisive, but I shall review the reasons for maintaining this assumption.

4.3.1 Somali is SIVO

Consider the sentences in (25) and (26).

(25)

a. naagtii way aragtay gabadhii
woman-the F-she saw girl-the
'The woman saw the girl'

b. naagtii gabadhii way aragtay
woman-the girl-the F-she saw
'The woman saw the girl' or
'The girl saw the woman'

(26)

a. ninkii wuu dilay shabeelkii
man-the F-he killed leopard-the
'The man killed the leopard'

b. ninkii shabeelkii wuu dilay
man-the leopard-the F-he killed
'The man killed the leopard' or
'The leopard killed the man'

In (25), both NPs are feminine singular; in (26), both NPs are masculine singular. Therefore, in each sentence both subject and object agree with the verb for person, number, and gender. In standard Somali, (25)b and (26)b would not be ambiguous because nominative Case marking results in high tone

on the determiner being lowered. However, in the dialect of one consultant, high tone is <u>not</u> lowered under nominative Case; for him, (25)b and (26)b are ambiguous, as indicated. On the basis of this evidence, we may say that (25)a and (26)a reflect the D-structure of Somali. In (25)b and (26)b, one NP has been preposed from post-verbal position; without nominative Case marking, the D-structure is unknown. (27) represents a partial S-structure for (25)a; (28) represents the ambiguous structures consistent with (26)b.

(27) naagtii way [$_{VP}$ aragtay gabadhii]

(28)

a. naagtii gabadhii$_i$ way [$_{VP}$ aragtay e$_i$]

b. naagtii$_i$ gabadhii way [$_{VP}$ aragtay e$_i$]

The second reason for holding Somali to be SIVO is theory-internal, based upon general assumptions regarding language acquisition. It has been proposed that languages may be parameterized with respect to the direction of branching in their major lexical categories, so that they are 'head first' or 'head last'. While the directionality of VP, the projection of the lexical category V, may be unclear, NP is unambiguously head first, as shown in (29).

(29) Saeed (1987, 111)

a. dal
country
'a country'

b. dalka
country-the
'the country'

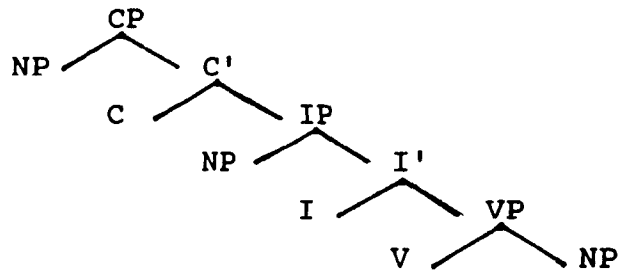
c. dalka weyn
country-the big
'The big country'

d. dalka aan tegayo
country-the I go
'the country I am going to'

If NP is head first, then in the absence of evidence to the contrary, it is a good hypothesis that VP is head first as well.[9]

I now summarize in (30) the various claims made to this point with respect to Somali phrase structure.

(30)



4.3.2 Arguments for SOVI

Before closing this discussion of Somali phrase structure, I want to review the evidence for SIOV word order. The strongest evidence for SIOV comes from nominative Case marking. The following argument and data are derived entirely from Livnat (1984, 62-71), to which the reader is referred for more details.

Nominative Case assignment marks only the right-most element of the subject. This is true of a simple noun, a noun modified by an adjective, or coordinate structures. In (31), the nominative Case morphology is underlined.

(31)

a. ninkū
man-the(+nom.)
'the man'

b. ninka fiicani
man-the good(+nom.)
'the good man'

c. ninka iyo naagtu
man-the and woman-the(+nom.)
'the man and the woman'

d. cuntada fiican ee kululli
food-the good and hot (+nom.)
'the good and hot food'

If NP contains a relative clause, Case is displayed through inflectional morphology on the verb. In (32)a-c, the bracketed NP contains a relative clause in which the object, bariiska, has been relativized.

(32) Livnat (1984, 67-70)

a. [bariiska [uu ninku iibiyaa]] wuu fiicanyahay
rice-the he man-the sells F-it is good
'The rice which the man sells is good.'

b. waan cunaa [bariiska [uu ninku iibiyo]]
F-I eat rice-the he man-the sells
'I eat the rice which the man sells.'

c. $[_{CP}$[bariiska [uu ninku iibiyo]]$_i$ baa $[_{IP}$ e_i fiican]]
rice-the he man-the sells F good
'The rice which the man sells is good.'

When the NP is subject of the sentence, it is marked with nominative Case, which appears as the inflectional morphology aa on the verb. When NP is not in subject position, because it is the object, as in (32)b, or because it is focused, as in (32)c, it receives non-nominative Case, which is realized as the inflectional morphology o on the verb.

Crucially, these distinctions are retained even if the verb is not the last word in the relative clause at surface structure, as in (33):

(33)

a. [bariiska [uu iibiy<u>aa</u> ninku]] wuu fiicanyahay
rice-the he sells man-the F-it is good
'The rice which the man sells is good.'
cf. (32)a

b. waan cunaa [bariiska [uu iibiy<u>o</u> ninku]]
F-I eat rice-the he sells man-the
'I eat the rice which the man sells.'
cf. (32)b

The contrasting verb morphology in (32)a-c and (33)a-b appears to indicate that the verb is the right-most element in the NP, and therefore in the relative clause. This would appear to be strong evidence that the clause in Somali is verb-final and that VP has the internal order OV. However, in the absence of a more complete analysis of the data, this evidence is not necessarily decisive.

The question of word order within VP must be regarded as unresolved, with reasonable arguments available for either SIVO or SIOV. In the absence of further evidence, I shall continue to assume that Somali is SIVO.

Notes

1. More precisely, waa only occurs with AGR_S, since AGR_W only occurs in the case of subject extraction--that is, with baa, ayaa, or the equivalent.

Note that it is unclear whether waa is generated under AGR or I. In the absence of any evidence that I am aware of, I am assuming the former.

2. This was brought to my attention by Mark Baker.

3. This claim will be qualified below.

4. The suggestion I investigate this approach is due to Mark Baker.

5. The analysis of FPs may extend to WH questions, as in (i):

(i) Livnat (1984, 109)

a. yaa qoray warqadda
who wrote letter-the
'Who wrote the letter?'

b. yuu arkay Cali
who-he saw Ali
'Who did Ali see?'

In (i)b, the subject Cali has been moved to post-verbal position, as indicated by the presence of the SCL uu.

6. For example, 'clitic climbing' is available in some languages, such as Italian:

i. Kayne (1989, 239)

Gianni li vuole vedere
John them wants to see
'John wants to see them'

7. It is unclear how to interprete apparent cliticization SCL onto in, as in (24)a; at this time, it appears we can say little more than that it is optional. Somali orthography may be misleading on this point. My consultants have been careful to pronounce the SCL in sentences like (24)a as a disinct word, indicating that (for them) it has not cliticized onto the complementizer.

8. One might argue that Somali is SOVI, on the assumption that the auxilary verb in sentences like i. to iii. from Saeed (1987, 33) are generated under I, on analogy with English auxiliaries.

i. waan keeni jiray
F-I bring used to
'I used to bring'

ii. waan keeni doonaa
F-I bring will
'I will bring'

iii. waan keeni lahaa
F-I bring would
'I would bring, I would have brought'

However, I think this assumption would be very dubious, and it will not be pursued further here.

9. It is often claimed in the traditional literature that the 'preferred' structure is SOV. It is not self-evident whether such judgments reflect D-structure word order or relatively late stylistic rules. One Somali consultant not only produced SVO word order, but explicitly reject SOV, calling it 'bad' Somali. There is no doubt that considerable dialectal variation may be found among Somali speakers, nor can interference from English word order be ruled out. Unfortunately, it is not easy to factor out these effects.

5. Subject Extraction and Strong Agreement

In the preceding section, I examined focus in Somali, with particular attention to how focus interacts with agreement processes and the SCL. For the sake of concreteness, the issues of word order within VP and of VP relative to I were addressed, if not answered conclusively. Although focus is a prominent feature of Somali syntax, the discussion took us somewhat afield. In this section, we shall return to discussion of subjects and agreement processes, and in particular to subject extraction.

5.1 Subject Extraction in Somali

In 3.2, it was proposed that a distinction is available in Somali between AGR_S and AGR_W. AGR_S selects verb morphology from the extensive paradigm, generates an optional SCL, and assigns nominative Case. AGR_W selects from the restrictive paradigm, does not generate a SCL, and does not assign nominative Case. AGR_W occurs with subject extraction from a matrix clause or by relativization; AGR_S occurs elsewhere. In particular, subject extraction from a sentential complement is from the domain of AGR_S, with an additional requirement that the SCL be present. The analyses of agreement processes and pro in earlier sections accounted for a significant range of data, but they do not account for these particular restrictions. Recall (10) from 3.2, repeated below as (1):

(1)

a.	nimanka	baa	dilay	shabeelka
	men-the(-nom.)	F	killed(r)	leopard-the
	'The men killed the leopard'			
b.	*nimanka	baa	<u>dileen</u>	shabeelka
	men-the(-nom.)	F	killed(e)	leopard-the
c.	*ninanka	<u>bay</u>	dilay	shabeelka
	men-the(-nom.)	F-they	killed(r)	leopard-the
d.	*nimank<u>u</u>	baa	dilay	shabeelka
	men-the(+nom.)	F	killed(r)	leopard-the
e.	*nimank<u>u</u>	<u>bay</u>	<u>dileen</u>	shabeelka
	men-the(+nom.)	F-they	killed(e)	leopard-the

The ungrammantical cases (1)b to (1)e contain elements that can only be generated under AGR_S. Take (1)e: on the basis of the present analysis, it has a S-structure like (2), irrelevant details aside.

(2) $*[_{CP}$ nimanku$_i$ $[_{C'}$ bay $[_{IP}$ e$_i$ AGR_S dileen shabeelka $]]]$

The subject has nominative Case, there is a SCL, and the verb displays morphology from the extensive paradigm.

Further examples are available from Livnat (1984). The following show that the SCL cannot occur if the subject has been focused.

(3) Livnat (1984, 57)

a. Cali baa yimid
Ali F came
'Ali came'

b. *Cali buu yimid
Ali F-he came
(same meaning)

(4) Livnat (1984, 57)

a. Cali baa keena bariiska
Ali F brings rice-the
'Ali brings the rice'

b. *Cali buu keena bariiska
Ali F-he brings rice-the
(same meaning)

The same facts obtain in cases of WH-extraction and relativization, as in (7)a and (9)a of 3.2, repeated below as (5) and (6).

(5) Wh-extraction

kumaa akhriyey buugga
who read(r) book-the
'Who read the book'

(6) Relativization

[inantii buugga akhrisay]
girl-the(-nom.) book-the read(r)

waxay araktay askari
what-she saw(r) soldier-the
'The girl who read the book saw the soldier'

(5) and (6) would be ungrammatical if they were to occur with the extensive paradigm and/or SCL, reflecting presence of AGR_S.

The generalization can be extended to a restriction in Somali against indefinite subject in a matrix clause with the FP w- (i.e. waa) and a verb. The following examples are from Saeed (1984, 170):

(7)

a. baabuur baa i dhaafay
truck F me passed
'A truck passed me'

b. *baabuur wuu i dhaafay
truck F-it me passed
(same meaning)

(8)

a. baabuurkii baa i dhaafay
truck-the F me passed
'The truck passed me'

b. baabuurkii wuu i dhaafay
truck-the F-it me passed
(same meaning)

AGR_W occurs only in the case of subject extraction, so the presence of the verb FP w- is diagnostic for AGR_S. Redundantly, SCL is also diagnostic for AGR_S. Indefinite NPs, like quantified expressions, are extracted at LF--either substituted for [NP,CP], or what is more likely, simply adjoined to IP (May 1985; Barriers, 5); the details of execution are irrelevant here. The extracted NP will have the status of an operator at LF binding a variable in subject position. Therefore, the ungrammaticality of (7)b is due to subject extraction from the domain of AGR_S; the difference is that subject extraction is non-overt and occurs at LF.[1]

As indicated above, the facts which obtain in the case of subject extraction from a sentential complement are quite different. In this case, AGR_S is obligatory, with the added requirement that the SCL be present, as in (9) to (11):

(9)

[ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ [$_{C'}$ in*(uu)
man-the F-I think that-he

[$_{IP}$ e$_i$ tegey shalay]]]]]
left yesterday

'I think that <u>the</u> <u>man</u> left yesterday'

(10)

[ardayda$_i$ ayuu [$_{CP}$ barruhu faray
students-they F-he teacher-the made

[$_{CP}$ [$_{C'}$ in*(ay) [$_{IP}$ e$_i$ buugga akhriyan]]]]]
that-they book-the read

'The teacher made <u>the</u> <u>students</u> read the book'

(11) Livnat (1984, 79)

[naagta$_i$ uu [$_{IP}$ Cali qabo [$_{CP}$ [$_{C'}$ in*(ay)
woman-the he Ali thinks that-she

[$_{IP}$ e$_i$ ninka buugga siissay]]]]] waa Amina
man-the book-the gave F Amina

'<u>The</u> <u>woman</u> that Ali thinks gave the book to the man is Amina'

In the sentential complement, the SCL and verb morphology from the extensive paradigm are diagnostic for AGR_S.[2]

Grammatical (9) to (11) contrast with ungrammatical (12), where subject has been extracted from the domain of AGR_W in the sentential complement.[3]

(12)

*[ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ [$_{C'}$ in
man-the F-I think that

[$_{IP}$ e$_i$ tegéy shalay]]]]]
left yesterday

'I think that <u>the</u> <u>man</u> left yesterday'

In this case, the verb is from the restrictive paradigm, which is diagnostic for AGR_W. As well, AGR_W does not generate a SCL or assign nominative Case.

Returning to the question of subject extraction from a matrix or relative clause, the prohibition against subject extraction from the domain of AGR_S may be expressed schematically as in (13), where CP is understood to be a simplex clause.

(13) $*[_{CP}\ O_i\ \ [_{IP}\ e_i\ \ AGR_S\ \ VP\ ]\ ]$

The extracted subject may be viewed as an operator (O) binding the empty subject position. The location and internal structure of VP is irrelevant.

It is highly unlikely that a restriction like (13) can be peculiar to Somali. In the first place, the theory is moving in the direction of assigning language-specific facts to the lexicon, and (13) is clearly a syntactic rule. Secondly, such a rule, if specific to Somali, would pose considerable problems for the language learner, particularly in the case of (7)b and (8)b, with non-overt movement. In fact, by looking at other null subject languages, namely Italian and Irish, we shall confirm that (13) is not peculiar to Somali.

Italian is a null subject language with strong subject-verb agreement. It has been known for some time that subject extraction is possible only from post-verbal position.[4] The argument from standard Italian is complex, however the evidence from the Trentino dialect is relatively clear. Recall (38) from 3.4, repeated below as (14):

(14) Trentino (Roberge 1986c, 122)

a. 'Na putela l' ei vegnuda
a girl she has come
'A girl came'

b. E vegnu 'na putela
has come a girl
(same meaning)

When the subject <u>'na putela</u> is in pre-verbal position, it is accompanied by an obligatory, co-referential SCL. However, in the case of free inversion, as in (14)b, the subject position is occupied by a pleonastic null pronominal. The verb agrees with the null subject. The SCL paradigm in Trentino is defective, lacking a 3s neuter SCL. Hence, the absence of SCL in (14)b.[5]

In the case of subject extraction, the SCL is not permitted, indicating that extraction has been from post-verbal position. This may be observed in (15) to (18). Notice that the same facts obtain in subject extraction from a simplex clause ((15) and (16)) and from a sentential complement ((17) and (18)).

(15) Jaeggli (1984, 136)

a. Quante putele è na via?
how-many girls are gone away
'How many girls have left?'

b. *Quante putele ele nade via?[6]

(16) Jaeggli (1984, 136)

a. Chi ha magnà?
Who has eaten
'Who has eaten?'

b. *Chi halc magnà?[7]

(17) Jaeggli (1984, 135)

a. Quante putele penset che sia vegnu?
how-many girls think-2s that are-3s come
'How many girls do you think have come?'

b. *Quante putele penset che le sia vegnude?

(18) Jaeggli (1984, 135)

a. Chi penset che magna?
Who think-2s that eat
'Who do you think is eating?'

b. *Chi penset che el magna?

There is no SCL in the grammatical examples, indicating that subject extraction has been effected from post-verbal position. I assume AGR_S is present in all null subject languages, i.e. languages with strong or 'rich' subject-verb agreement. Consider the S-structure for ungrammatical (16)b:

(19) *$[_{CP}$ chi_i $[_{IP}$ e_i AGR_S halo magnà]]

The structure of (19) is identical, in all relevant respects, to (13) above, the ungrammatical structure found in Somali. The difference between Somali and Italian is in the device used to circumvent (13). In Somali, it is a distinction between AGR_S and AGR_W; in Italian, it is free inversion of subject and verb.

Now consider Irish, a null subject language, which differs in important ways from both Somali and Trentino, yet shares the prohibition captured in (13). With respect to Irish verb inflection, McCloskey and Hale (1984, henceforth M & H, 489) report:

> Verbal paradigms consist of two kinds of forms for which we will use the traditional terms 'analytic' and 'synthetic' respectively. More precisely, each verbal paradigm consists of exactly one analytic form and a set, possibly null, of synthetic forms. The synthetic form represents an inflectional type familiar from many other Indo-European languages. In a single inflectional ending it encodes information about tense and mood, as well as the person and number of its subject. The analytic form encodes only information about tense and mood, but not about person-number characteristics of its subject.

M & H (p. 489) supply the paradigm of the conditional of the verb <u>cuir</u> 'put' in the Ulster dialect:

(20)	1s	chuirfin		1p	chuirfimis
	2s	chuirfeá		2p	chuirfeadh sibh you(pl)
	3sm	chuirfeadh	sé he	3p	chuirfeadh siad they
	3sf	chuirfeadh	sí she		

The synthetic forms are 1s, 2s, and 1p; the analytic form is <u>chuirfeadh</u>. If a synthetic form does not exist, as in the 2p and all third person forms for the conditional of <u>cuir</u>, the analytic form is used, along with an independent subject pronoun.

The analytic form has the following distribution:

a) with independent subject pronouns, when a corresponding synthetic form is not available (e.g. 2p, 3s, and 3p in (20)

b) with overt lexical NP subject

c) with trace subject, even if an appropriate synthetic form is available, as in (21)

(21) Chan mise a chuirfeadh t isteach ar
Cop+Neg me that put(CONDIT) in on

an phost sin
that job
'It's not me that would apply for that job'

M & H (p. 490) report, with respect to (21) (their (5)):

> The antecedent for the subject trace in (5) is a S1 pronoun. Note however, that even though there exists a synthetic form of <u>cuir</u> appropriate for the context of (5), one finds instead the analytic form <u>chuirfeadh</u>. This is an obligatory effect. That is, use of the S1 form <u>chuirfinn</u> in a context like that in (5) is ungrammatical. Summarizing this distributional pattern, we can say that if the subject of a clause is phonologically specified, then the verb of that clause may not carry person-number inflection. Furthermore, wh-traces pattern with lexical NP with respect to this effect.

We can see that Irish has a contrast between synthetic and analytic forms which corresponds generally to the distinction between AGR_S and AGR_W, and all that falls out from that distinction. However, Somali has 'paired' verb inflection paradigms, with an entire extensive paradigm, corresponding to AGR_S, and a restrictive paradigm, corresponding to AGR_W. Irish is a mixed verb inflection system, with synthetic forms (corresponding to AGR_S) available for some, but not all, person-numbers of a particular verb paradigm and a completely leveled weak paradigm, corresponding to AGR_W. The crucial facts for present purposes are illustrated in (21). In the case of subject extraction, the analytic form (corresponding to AGR_W) is required, even if a synthetic form (corresponding to AGR_S) is available, and otherwise would be used.

The relevant facts from these languages are summarized in (22):

(22) Possible occurence in the domain of AGR_S

	pro	full pronoun	lexical NP	WH-trace
Somali	+	+	+	-
Trentino[8]	+	?	+	-
Irish	+	-	-	-

M & H (p. 490) note that WH-trace patterns with lexical NP. However, on the basis of (22), it seems more accurate to say that in Irish lexical NP patterns with trace.

I began by saying that there are theoretical reasons why (13), repeated below, is not peculiar to Somali.

(13) $*[_{CP}\ O_i\ [_{IP}\ e_i\ AGR_S\ VP\ ]\ ]$

Now we have seen evidence from Trentino and Irish indicating that (13) is valid for these null subject languages as well. As (13), evidently, can be attributed to the human language faculty, we need to determine its theoretical status. In particular, we should expect (13) to follow from general linguistic principles.

At this point, it will be convenient to summarize the Somali data which must be accounted for. First, there is the prohibition against short subject extraction from the domain of AGR_S in matrix or relative clause, as illustrated by (1)a and (1)e, repeated here as (23):

(23)

a. nimanka baa dilay shabeelka
men-the(-nom.) F killed(r) leopard-the
'The men killed the leopard'

b. *nimanku bay dileen shabeelka
men-the(+nom.) F-they killed(e) leopard-the

Secondly, subject extraction from a sentential complement is from the domain of AGR_S, with additional requirement that SCL be present. Subject extraction from the domain of AGR_W is not possible in this configuration. See (12) and (9), repeated here as (24):

(24)

a. *[ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ [$_{C'}$ in
man-the F-I think that

[$_{IP}$ e$_i$ tegéy shalay]]]]]
left yesterday

'I think that <u>the</u> <u>man</u> left yesterday'

b. [ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ [$_{C'}$ in*(uu)
man-the F-I think that-he

[$_{IP}$ e$_i$ tegey shalay]]]]]
left yesterday

(same meaning)

(23)a is an ordinary instance of grammatical subject extraction, a common and familiar phenomenon. (24)a is also relatively unproblematic. We know that AGR_W cannot identify pro. Therefore, the empty subject position can only be an A'-bound trace, which must be properly governed and antecedent governed by an intermediate trace in [NP,CP]. Evidently, the complementizer <u>in</u> blocks government. These are the familiar that-t or ECP effects. The approach suggested in <u>Barriers</u> invokes the Minimality Condition, such that the complementizer governs the subject position and blocks antecedent government by a more distant intermediate trace. Hence, (24)a may be assimilated to these more familiar cases. The problematic cases are (23)b and (24)b; the several approaches we consider will none of them differ significantly in the treatment of (23)a or (24)a.

5.2 Previous Treatments of Subject Extraction in Null Subject Languages

The problem that we are examinig is not a new one, although its status, even the definition of what the problem is, has changed as work in generative grammar has progressed. The earliest discussion of subject extraction in null subject languages was Perlmutter (1971), who observed that languages like Italian do not exhibit that-t effects. Rizzi (1982, c.4, however the paper was circulated informally as early as 1980) demonstrated that subject extraction in Italian was from post-verbal position. Chomsky (1981), in developing a new theoretical framework, linked Rizzi's results with the ECP--free inversion and subject extraction from post-verbal position permitted apparent ECP violations in Italian. The work by Rizzi and Chomsky focused on 'long' extraction of a subject from post-verbal position in an embedded clause, and assumed that direct subject extraction from this position was blocked by that-t effects, or the ECP, depending upon the current version of the theory. Rizzi (1982, 151-3) did, however, recognize that subject extraction from a matrix clause also occurs from post-verbal position.

I shall not review the analyses presented in this earlier work--they are well-known and would take us too far afield. I shall restrict myself to reviewing two more recent approaches which address the problem of short subject extraction in null subject languages, a configuration where appeal to the ECP seems inadequate.

5.2.1 Jaeggli's Approach

Jaeggli (1984) takes as his starting point the 'Concepts and Consequences' version of government binding theory (Chomsky, 1982). A number of elements in this framework are

crucial to his analysis. First, binding theory partitions the class of NPs into four categories:

(25)

a. [+anaphor, -pronominal]

b. [-anaphor, +pronominal]

c. [+anaphor, +pronominal]

d. [-anaphor, -pronominal]

Corresponding to this typology are four kinds of empty category: NP-trace, pro, PRO, and variable. The second of these, pro, was introduced in Chomsky (1982) and corresponds to the null subject in a null subject language. A special requirement applies to pro, in that it must be identified by 'rich' agreement (Chomsky 1982, 78-89). Jaeggli proposes that pro is 'identified' through government:

(26) An empty category is [-anaphor, +pronominal] (i.e. 'pro'--APH) iff it is governed by 'rich' agreement.

Note that in (26), Jaeggli accepts the functional determination of empty categories. Suppose, then that the subject of (27)a is extracted directly from subject position, yielding the S-structure (27)b.

(27)

a. Chi ha comprato una casa
Who has bought a house

b. *[Chi_i $[_{IP}$ e_i AGR_S ha comprato una casa]]

In (27)b, having a structure identical to (13), the empty category in subject position is governed by AGR_S (= 'rich' agreement), which determines that the empty category is pro. The next step is to prevent a pronominal from being locally

bound by an operator. (27)b will then be ruled ungrammatical as an instance of vacuous quantification.

Jaeggli notes, however, that a pronoun can be given a bound interpretation under some circumstances. Consider the Spanish examples (28) and (29):

(28) Quienes piensan que ellos son inteligentes
Who thinks that they are intelligent

(29) Quienes dijeron que e piensan que ellos son inteligentes
Who said that think-3pl that they are intelligent

In (28), ellos cannot be given a bound interpretation; but in (28), it can. Jaeggli draws on Montalbetti's observation that (in Spanish) overt pronouns cannot be bound, unless linked to a bound pronominal. Under this interpretation, ellos in (29) is linked to e, a pro which is itself bound by the trace of quienes. The conclusion drawn is that a pronominal can be bound via a trace, that is, a pronoun can be bound in a structure with an LF representation like (30).

(30) $Q_i \ldots x_i \ldots$ pronominal$_i$

This observation motivates Jaeggli's Condition on Bound Pronominals:

(31) Condition on Bound Pronominals[9]

A pronominal p can be interpreted as a variable bound by a quantifier Q only if bound by a variable of Q.

Consider now how (31) would apply to the Somali data in (23)b and (24)b:

(23)b

*nimanku bay dileen shabeelka
men-the(+nom.) F-they killed(e) leopard-the
'The men killed the leopard'

*[nimanku$_i$ bay [$_{IP}$ e$_i$ AGR$_S$ dileen shabeelka]]]

(24)b

ninka baan ramaysanahay inuu tegey
man-the F-I think that-he left
'I think that the man left'

[ninka$_i$ baan [$_{IP}$ ramaysanahay
man-the F-I think

[$_{CP}$ t$_i$ [$_{C'}$ inuu [$_{IP}$ e$_i$ AGR$_S$ tegey]]]
that-he left

In (23)b, the presence of AGR$_S$ identifies the subject as a pro. The Condition on Bound Pronominals is violated, and the sentence is predicted to be ungrammatical, which it is.

In (24)b, AGR$_S$ likewise identifies the subject as pro, which is bound by an operator, <u>ninka</u>, and presumably by an intermediate trace in [NP,CP] of the sentential complement. However, the intermediate trace is not a variable, since it is in an A' position. Therefore, the Condition on Bound Pronominals predicts (24)b to be ungrammatical, which is false. (24)b is fully grammatical.

The approach of Jaeggli (1984), aside from being expressed in a framework that is no longer current, fails to make the correct predictions with respect to long subject extraction in Somali, as in (24)b. As well, it introduces a new condition with relatively narrow empirical implications.

Let us now consider the approach taken by Y. Roberge.

5.2.2 Roberge's Approach

The orientation of Roberge (1986c) is different from the one pursued here, as he is particularly concerned with clitic phenomena, including clitic doubling. He proposes a definition of C (= clitic) chain:

(32) C-chain (definition) (Roberge 1986c, 225)

a and b form a C-chain if:
i. a and b are co-indexed;
ii. a governs b;
iii. b is locally free[10]

Conditions (32)i and (32)ii have the effect of licensing pro in the argument position (= b) associated with a clitic (= a). Roberge states, 'The last condition in (iii) is new and is introduced in order to block any extraction out of the argument position (b) related to a clitic (a) since b would otherwise be locally A'-bound by the coindexed antecedent.' (Roberge 1986c, 225). To (32), Roberge adds (Roberge 1986c, 226) condition (33):

(33) A clitic must be part of a C-chain.

Between them, (32) and (33) prevent extraction of a clitic-doubled NP and prevent the occurence of an overt or non-overt anaphor in a clitic-doubled context. In either case, the clitic must be part of a C-chain by (33), and therefore (32) must be satisfied. If a clitic-doubled NP is extracted, it will A'-bind its trace, and the sentence will be out by (32)iii. If an anaphor is clitic-doubled, it will be locally A-bound, and again the sentence will be out by (32)iii.[11]

If (33) applies at LF (assumed to be the unmarked case), extraction from a clitic-doubled position is blocked absolutely. If it applies at S-structure, it blocks overt extraction, e.g. WH-movement, but permits clitic-doubled

quantified expressions, which undergo QR at LF. (This evidently occurs in Trentino, Fiorentino, and Pied Noir French.) Finally, if (33) does not apply at all, extraction from clitic-doubled positions will be freely permitted, as evidently occurs in Argentinian and Uruguayan Spanish. Roberge (1986c) correctly predicts that no language will permit overt extraction of a clitic-doubled NP at S-structure, but exclude clitic-doubled quantified expressions.

As Roberge indicates (1986c, 225), his approach can be extended readily to the licensing of pro by agreement, i.e. by AGR_S. Applying his approach to (13), it might be argued that AGR_S, parallel to a clitic, must be part of a chain co-indexed with [NP,IP]. Consider how this might apply to the Somali data in (23)b and (24)b. In (23)b, AGR_S governs and by convention is co-indexed with the empty subject position. However, the empty subject position is bound by nimanku, so (23)b violates (32)iii and is ungrammatical, as predicted.

I have already assimilated (24)a to the familiar that-t effects. By the Minimality Condition, the complementizer in blocks antecedent government of the subject position. Consequently the trace in subject position is not properly governed, resulting in an ECP violation.

(24)b is more problematic. The empty subject position in the sentential complement is governed by and co-indexed with AGR_S. We may assume an intermediate trace in [NP,CP] of the sentential complement, which m-commands and presumably binds the empty subject position. The Minimality Condition is relevant to government, but it is not relevant to binding. In (24)b, the Minimality Condition does not block binding of the subject position, (24)b violates (32)iii and should be ungrammatical, which it is not.

Suppose we permit deletion of the intermediate trace, which is permissible if the empty subject is a pronominal, i.e. not a trace. As (32)i and (32)ii license pro in subject

position, the approach would appear to be tenable. In this case, the empty subject and its antecedent will be co-indexed by construal, since they do not form a chain.[12]

Subject to these qualifications, Roberge (1986c) obtains the correct results. However (32)iii. is problematic. It is redundant with respect to lexical NPs and pronouns, as they must be free and locally free, respectively, by the binding theory. Therefore, the only structure with empirical consequences is trace forming a C-chain with a clitic or agreement features. This is precisely the structure (13), which I have observed to be ungrammatical. Therefore (32)iii is essentially a stipulation and does not answer the more fundamental question of why this structure is ungrammatical. It is this question that I shall address in the next section.

5.3 Subject Extraction from the Domain of AGR_S

The problem is to account for the fact that short subject extraction in Somali from matrix or relative clause is only possible from the domain of AGR_W, as in (23), while long subject extraction from sentential complement is only possible from the domain of AGR_S, with the additional requirement in the latter case that the SCL be present in the sentential complement. (23) is repeated below:

(23)

a. nimanka baa dilay shabeelka
men-the(-nom.) F killed(r) leopard-the
'The men killed the leopard'

b. *nimanku bay dileen shabeelka
men-the(+nom.) F-they killed(e) leopard-the

(23)a is a normal, grammatical case of short subject extraction. What is problematic is the prohibition against subject extraction from the domain of AGR_S, as in (23)b.

Building on the insights of Jaeggli (1984), the approach I shall take is to attribute the ungrammaticality of (23)b to a binding theory violation. More specifically, if the empty subject of (23)b is pro, and if the governing category is CP, then (23)b will violate Principle B of the binding theory, which requires pronominals to be free in their governing category.

The next step is to accomodate the contrast between licit and illicit A'-binding as in (23) to the binding theory. The binding theory of Chomsky (1981) is a theory of A-binding and will not make the proper distinctions. Therefore, I shall adopt the generalized binding theory of Aoun (1985), which is a theory of A- and A'-binding:

(34) Generalized Binding (Aoun 1985, 28)

A. An anaphor must be X-bound in its governing category

B. A pronominal must be X-free in its governing category

C. A name must be A-free

(Where X=A or A')

I shall define governing category then as (35):

(35) Governing Category (def)

<u>a</u> is the governing category for <u>b</u> iff <u>a</u> is the minimal NP or CP containing <u>b</u>, a governor of <u>b</u>, and a SUBJECT accessible to <u>b</u>.

Returning to (23)b, I have argued that the empty subject position is pro. By (35), the governing category for pro is CP, in which it is A'-bound by the antecedent <u>nimanku</u>, violating Principle B of (34). It follows that (23)b is ungrammatical.

Note that sentences like (23)b are ungrammatical even when SCL is absent. (See (1)a-e above.) This follows from

the proposal in 3.3, based upon Rizzi (1986), that AGR_S licenses pro, while SCL and the extensive verb paradigm give it content. The presence of AGR_S, as indicated for instance by the extensive verb paradigm, is enough to identify a null subject to be a pronominal for purposes of the binding theory, even if SCL is absent and pro therefore lacks content.

Now consider long extraction from the subject position of a sentential complement, as in (24), repeated below:

(24)

a. *[ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ t$_i$ [$_{C'}$ in
man-the F-I think that

[$_{IP}$ e$_i$ tegey shalay]]]]]
left yesterday

'I think that <u>the</u> <u>man</u> left yesterday'

b. [ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ t$_i$ [$_{C'}$ in*(uu)
man-the F-I think that-he

[$_{IP}$ e$_i$ tegey shalay]]]]]
left yesterday

(same meaning)

It was indicated above that (24)a is a standard case of that-t or ECP effects and requires no further comment.[13]

Now consider (24)b. The null subject is pro, because it is in the domain of AGR_S. Observe that the intermediate trace binds pro within CP, its governing category. Therefore, it appears that (24)b should represent a Principle B binding theory violation. Nevertheless, the sentence is fully grammatical.

Insofar as it entails these results, the representation (24)b must be incorrect. Let us consider, therefore, an approach suggested in 5.2.2, namely that the intermediate trace in [NP,CP] may be deleted before application of binding theory. There is no principled reason why this cannot occur,

since pro does not require an intervening trace for its interpretation. The derivation will then proceed as follows: at S-structure, the subject will be extracted via movement to the embedded [NP,CP], adjunction to VP, and movement to [NP,CP] of the matrix clause, leaving an empty category at the extraction site which is licensed as pro by AGR_S and given content conjointly by the extensive paradigm verb morphology and the SCL. The intermediate trace in [NP,CP] will delete before application of binding theory at LF, insuring that pro is free in its governing category, as required. The antecedent and null subject may retain the same index, or they may be co-indexed by construal; failure of co-indexation will result in a case of vacuous quantification. The resulting structure will be like (25). (I assume the trace adjoined to VP is deleted as well, though the question is immaterial here.)[14]

There is independent evidence to support deletion of intermediate trace. A lexical subject generated in the domain of AGR_S is assigned nominative Case. This is true whether the subject is in D-structure position or is clause-bound moved to another position, as in (22) of 3.3, repeated here as (36):

(36) Livnat (1984, 10)

a. Cali wuxuu moodayey [inay <u>nimanku</u> tageen]
Ali F-he thought that-they men-the(+nom.) left
'Ali thought that the men had left'

b. Cali wuxuu moodayey <u>nimanku</u>$_i$ [inay e_i tageen]
Ali F-he thought men-the(+nom.) that-they left
(same meaning)

If the subject of a simplex clause is extracted, it displays non-nominative Case, as in (1)a. This is due to being generated within the domain of AGR_W, which I have already argued does not assign nominative Case. The position of [NP,CP] does not exclude nominative Case intrinsically. Normally, an extracted NP forms a chain with the extraction

site; Case is assigned to the extraction site and is displayed on the NP with which it forms a chain, however the details are accounted for.[15]

Now, compare (36) with cases of long subject extraction from a sentential complement, as in (9) to (11), repeated below:

(9)

[ninka$_i$ baan [$_{IP}$ e ramaysanay [$_{CP}$ t$_i$ [$_{C'}$ in*(uu)
man-the F-I think that-he

[$_{IP}$ e$_i$ tegey shalay]]]]]
left yesterday

'I think that <u>the</u> <u>man</u> left yesterday'

(10)

[ardayda$_i$ ayuu [$_{CP}$ barruhu faray
students-they F-he teacher-the made

[$_{CP}$ t$_i$ [$_{C'}$ in*(ay) [$_{IP}$ e$_i$ AGR$_S$ buugga akhriyan]]]]]
that-they book-the read

'The teacher made <u>the</u> <u>students</u> read the book'

(11) Livnat (1984, 79)

[naagta$_i$ uu [$_{IP}$ Cali qabo [$_{CP}$ t$_i$ [$_{C'}$ in*(ay)
woman-the he Ali thinks that-she

[$_{IP}$ e$_i$ ninka buugga siissay]]]]] waa Amina
man-the book-the gave F Amina

'<u>The</u> <u>woman</u> that Ali thinks gave the book to the man is Amina'

In (9) to (11), subject extraction has been from the domain of AGR$_S$. Therefore, we should expect the extracted NP to display nominative Case. Surprisingly, they are non-nominative. This may be explained, however, given the present analysis. Unlike clause-bound NP movement, which occurs at PF, extraction occurs at S-structure, leaving an intermediate

trace in [NP,CP] of the sentential complement. If this intermediate trace is deleted, as I claim, the antecendent and pro will not form a chain. Hence, if nominative Case is assigned at PF, the chain terminating in the extraction site will consist of one member, namely pro, and nominative Case will not be transported to the subject. What appears on the subject is non-nominative Case, the default Case, assigned under what are essentially 'elsewhere' conditions.[16]

To summarize the argument, the analysis of short subject extraction in Somali from matrix or relative clause has been based upon the assumption that AGR_S determines a null subject to be pro. Ungrammatical short subject extraction from the domain of AGR_S then reduces to a Principle B binding theory violation. With respect to long subject extraction from sentential complement, complementizer blocks antecedent government by an intermediate trace, so that subject extraction from the domain of AGR_W would result in an ECP violation. Valid long extraction leaves pro, which is licensed by AGR_S and given content through the presence of extensive verb paradigm morphology and SCL. Subsequent deletion of the intermediate trace prevents a Principle B violation.

In this section, I assimilated a range of Somali data to a general prohibition against direct subject extraction from the domain of AGR_S. I then made a series of proposals, based chiefly upon the Somali data, to account for this prohibition. It remains now to demonstrate that these proposals are of general validity, and in particular that they may be extended to phenomena in other languages. In the best case, we should hope to find an improved explanation for what had been problematic. At the least, we should not create more problems than we found. I shall now return to the cases of Italian and Irish, to demonstrate the effects of the present proposals for these null subject languages.

5.4 Extending the Account to Italian and Irish

Recall that subject extraction in Italian, including the Trentino dialect, is from post-verbal position. This is true of subject extraction from a simplex clause, as in (15) and (16), repeated below, and from a sentential complement, as in (17) and (18):

(15) Jaeggli (1984, 136)

a. Quante putele è na via?
how-many girls are gone away
'How many girls have left?'

b. *Quante putele ele nade via?

(16) Jaeggli (1984, 136)

a. Chi ha magnà?
Who has eaten
'Who has eaten?'

b. *Chi halo magnà?

(17) Jaeggli (1984, 135)

a. Quante putele penset che sia vegnu?
how-many girls think-2s that are-3s come
'How many girls do you think have come?'

b. *Quante putele penset che le sia vegnude?

(18) Jaeggli (1984, 135)

a. Chi penset che magna?
Who think-2s that eat
'Who do you think is eating?'

b. *Chi penset che el magna?

The present account may be applied in a straight-forward manner to the simplex clause, as in (15) and (16). the presence of SCL is diagnostic for direct subject extraction,

which I have argued violates Principle B of the binding theory. Subject extraction from post-verbal position avoids this prohibition on the assumption that the post-verbal extraction site is occupied by a trace and that the subject position is occupied by a null pleonastic pronominal, which is not bound by the extracted NP. On the basis of only (15) and (16), subject extraction from post-verbal position might be interpreted as a device available in Italian for preventing a Principle B binding theory violation.

The story is quite different for subject extraction from sentential complement, as in (17) and (18). Direct subject extraction would be structurally identical to the Somali case: the null subject at the extraction site should be licensed and identified as pro, and the intermediate trace should delete, permitting the derivation. And yet, direct subject extraction from a sentential complement clearly does not occur.

The most promising response to this problem is to attribute post-verbal subject extraction to some independent feature or features of Italian. In fact, this is a plausible approach, since Italian, but not Somali, permits free inversion, as discussed in section 3.4, indicating that the relation of subject to I and V in these two languages must differ at some fundamental level.

One possibility, consistent with this observation, is that the subject is generated post-verbally in Italian. Adams (1987) is only one of a number of authors to make such a proposal.[17] If this approach is tenable, then 'free-inversion' is accounted for directly, as is subject extraction from post-verbal position. It is beyond the scope of the present study to investigate this approach to Italian. For our purposes, the availability of such an account undermines the relevance of (17) and (18) as counter-evidence to the proposals made in 5.3.

In summary, the evidence from Italian is relatively neutral with respect to the arguments made above. Subject

extraction from a simplex clause is accounted for, and subject extraction from sentential complement is inconclusive as counter-evidence.

Compared to Italian, the data from Irish is more complex and more interesting.

It was indicated in 5.1 that Irish, like Somali, has a contrast between strong and weak subject-verb agreement (AGR_S and AGR_W), corresponding to the 'synthetic' and 'analytic' verb paradigms. Synthetic forms incorporate person-number features, as well as tense and mood; analytic forms incorporate tense and mood only. As indicated in (22), analytic forms occur with WH-trace, as well as with lexical NPs and lexical pronouns; the synthetic forms occur only with pro. The crucial evidence for subject extraction from the domain of AGR_W was presented in (21), repeated below. Note that the analytic form <u>chuirfeadh</u> (corresponding to AGR_W) occurs, rather than the synthetic form <u>cuir</u> (corresponding to AGR_S), which is available and would otherwise occur.

(21) Chan mise a chuirfeadh t isteach ar
Cop+Neg me that put(CONDIT) in on

an phost sin
that job
'It's not me that would apply for that job'

The facts in Irish are actually a good deal more complex than this. McCloskey (1979) discusses two relative clause strategies available in Irish, for which he uses the traditional terms 'direct' and 'indirect' relatives. The direct relative leaves a variable (a 'gap' in McCloskey's terminology), while the indirect relative leaves a resumptive pronoun. In a simplex clause, the distribution is as follows: direct relative is obligatory if the subject is relativized, direct and indirect relative are available if the direct object is relativized,[18] and the indirect relative is

obligatory for relativization of more remote NP, e.g. object of a preposition or NP specifier of a NP.

Corresponding to these strategies are two distinct complementizers, which McCloskey represents respectively as aL (a particle which appears as a and induces lenition of the following segment) and aN (a particle which also appears as a, but induces nasalization of the following segment). aL occurs with direct relatives, and aN occurs with indirect relatives. Therefore, these complementizers are diagnostic for the respective relativization strategies.[19] Examples of the two relativization strategies are given below:

(37) McCloskey (1979, 12)

a. Direct relative

an fear aL thuigeann e an scéal
the man that understands the story
'the man that understands the story'

b. Indirect relative

an fear aN n-insíonn tu an scéal dó
the man that tell you the story to him
'the man that you told the story to'

Following the line of analysis taken to this point, I am assuming that the direct and indirect relative structures in a simplex clause are like (38):

(38)

a. Direct relative

[NP [$_{CP}$ O_i [$_{C'}$ aL [$_{IP}$ ...t_i ...]]]]

b. Indirect relative

[NP [$_{CP}$ O_i [$_{C'}$ aN [$_{IP}$ ...pronoun$_i$ ..]]]]

A null operator generated in argument position will move to [NP,CP] at S-structure, leaving a variable in the case of

the direct relative and a bound pronominal in the case of the indirect relative.

Subject extraction, as in (37)a, is only possible from the domain of AGR_W, and using the direct relative strategy. AGR_W does not license and identify a null subject as pro, so the subject position must be a variable. AGR_W governs the subject position, but does not properly govern it, so the variable must be antecedent governed. In the ordinary case, a complementizer blocks antecedent government, presumably by the Minimality Condition. However, this is evidently not true of (37)a, having a structure similar to (38)a. We may instantiate this observation with ɲ the present framework by assuming that in the unmarked case, a complementizer is a governor and therefore invokes the Minimality Condition when a subject is extracted over it to the position [NP,CP]. It then follows that aL is marked as a non-governor, so it does not invoke the Minimality Condition. This should pose no problem for the language learner, as there is positive evidence for the marked status of aL. Informally, we may say that the marked status of aL is functionally equivalent to complementizer deletion in English.

We may take aN to be the unmarked case, a complementizer which is a governor and which therefore invokes the Minimality Condition. A subject cannot be extracted from the domain of AGR_W over aN, because it will leave a variable that is not properly governed. In a simplex clause, a subject cannot be extracted from the domain of AGR_S over aN, because the null subject will be licensed as pro and will be bound in its governing category, a Principle B binding theory violation.[20]

Next, let us consider extraction from embedded clauses, in which direct and indirect relative strategies have the partial structures (39)a-b, derived from McCloskey (1979) and M & H:

(39)

a. Direct relative

NP $[_{CP}$ O_i $[_{C'}$ aL $[_{IP}$ ... $[_{CP}$ $[_{C'}$ aL $[_{IP}$ ... t_i ...

b. Indirect relative

NP $[_{CP}$ O_i $[_{C'}$ aN $[_{IP}$ ... $[_{CP}$ $[_{C'}$ goN $[_{IP}$ ... pronominal$_i$ ...

In the direct relative, every clause is headed by aL. In the indirect relative, the highest clause is headed by aN, and every embedded clause is headed by goN, the complementizer which heads an ordinary sentential complement.

In the case of subject extraction from sentential complement, three strategies are available, which reduce in part to two. The direct relative may be used, in which case the subject is extracted from the domain of AGR_W and the complementizer aL appears at the head of each clause. Or, the indirect relative may be used, in which case a pronominal is left at the extraction site, and aN heads the highest clause, while each embedded clause is headed by goN. The pronominal may be a lexical pronoun, in which case the verb is of the analytic form, reflecting AGR_W. Or, the pronominal may be pro, if there is an available synthetic form, reflecting AGR_S. These variants are illustrated in (40) to (42), from M & H (498-9):

(40) Direct relative

na daoine aL bhí mé ag dúil aL chuirfeadh t
the people that was I expect (PROG) that put(CONDIT)

isteach ar an phost sin
in on that job
'the people that I was expecting would apply for that job'

(41) Indirect relative with lexical pronoun

na daoine aN raibh mé ag dúil goN gcuirfeadh
the people that was I expect (PROG) that put(CONDIT)

siad isteach ar an phost sin
they in on that job
'the people that I expected (that they) would apply for that job'

(42) Indirect relative with pro

na daoine aN raibh mé ag dúil goN
the people that was I expect (PROG) that

gcuirfidis pro isteach ar an phost sin
put(CONDIT P3) in on that job
'the people that I expected (that they) would apply for that job'

The analysis of the direct relative in (40) carries over directly from the analysis of direct relative in the simplex clause. The structure of the complementizer position of the embedded clause in (40) will be essentially as in (43):

(43) $\ldots[_{CP}\ t_i\ [_{C'}\ aL\ [_{IP}\ \ldots t_i \ldots AGR_W\ \ldots$

The subject is extracted from the domain of AGR_W over the complementizer aL, which does not block antecedent government. The extraction site is occupied by a variable which is bound in its governing category CP, as required under Principle A of the generalized binding theory. The intermediate trace cannot be deleted.

Now consider the indirect relative, which will have an internal structure like (44):

(44) $\ldots[_{CP}\ (t)\ [_{C'}\ goN\ [_{IP}\ \ldots\ pronominal_i\ \ldots AGR\ldots$

If the embedded verb displays morphology from the synthetic paradigm, reflecting AGR_S, the pronominal is pro. If the embedded verb is analytic, reflecting AGR_W, the pronominal is a full lexical pronoun. In either case, goN

blocks antecedent government, but does not block binding, so the intermediate trace (enclosed in parenthesis in (44)), must delete before application of binding theory.[21]

Subject extraction in Irish is quite complex, and yet it is consistent with the proposals in 5.3. The marked status of the complementizer aL permits subject extraction from the domain of AGR_W in a sentential complement. On the other hand, subject extraction by the indirect strategy over the complementizer goN requires a resumptive pronoun. The proposals in 5.3 carry over directly to this configuration. Irish differs from Somali only in permitting either pro or lexical pronoun in this configuration.

Irish provides a rich array of data respecting subject extraction, with important distinctions made with respect to agreement processes, resumptive pronouns, and complementizers. The fact that the proposals in 5.3 can be extended in a natural way to the Irish data indicates that these proposals are of some general validity.

Notes

1. The prohibition against subject extraction as in (5)b is widely attested. Unfortunately, I cannot comment on the grammaticality of (i):

(i) $[_{CP}\ NP_i\ \text{baa}\ [_{IP}\ NP_{[-def]}\ V\ e_i\ ]\ ]$

That is, indefinite subject is <u>in situ</u> and another NP is focused.

2. Note that the extracted subjects <u>ninka</u> in (10), <u>ardayda</u> in (11), and <u>naagta</u> in (12) display (surprisingly) non-nominative Case, although they have been extracted from the domain of AGR_S. I shall account for this in 5.3.

3. (12) has not been obtained from or verified with a native speaker.

Somali has a complex verb system with distinctions made between matrix clause, subordinate clause in a subject, and subordinate clause in a non-subject argument of the matrix verb. Corresponding to each of these are 'extensive' and 'restrictive' forms, for a total of six paradigms, although several of these are identical. (See note 4, section 2 for examples.) The restrictive paradigm does not occur in sentential complements, but it does occur in subordinate clauses of other non-subject arguments of the matrix verb.

The form of this verb in the simple past was obtained from Saeed (1987, 60-62).

4. See Rizzi (1982, c.4) and subsequent work by Rizzi and others.

5. Jaeggli and Roberge differ slightly in their accounts of SCLs in Trentino. Jaeggli (1984) claims that they are absent (i.e. in some general sense) in the case of free inversion. Roberge (1986c) attributes the absence of SCL in free inversion to the defective nature of the Trentino paradigm, which lacks a 3s neuter SCL. As Roberge's account is consistent with Fiorentino (see 3.4), I assume it is correct. The question is immaterial for the present discussion.

6. The 3sf SCL <u>le</u> is cliticized onto the verb <u>e</u>, yielding <u>ele</u>.

7. The 3sm SCL <u>lo</u> is cliticized onto <u>ha</u>, yielding <u>halo</u>.

8. I have no information on the possibility of full subject pronouns in Trentino.

9. Jaeggli proposes a revised version of (24) which as slightly different empirical coverage. (24) suffices for present purposes.

10. By 'locally free', Roberge evidently means 'free in its governing category.

11. Sentences with the French reflexive clitic se present an apparent problem for Roberge's analysis, as in (i):

(i) Jean se voit.
'Jean sees himself.'

See Roberge (1986c, 243-251) for discussion.

12. In fact, I shall incorporate deletion of intermediate trace into my own analysis below.

13. Aoun's generalized binding had as one of its consequences the elimination of the ECP as an independent principle. Within the pre-Barriers model, an antecedent in COMP would not c-command its trace in subject position if the complementizer was present (i.e. there was a 'doubly-filled' COMP), and consequently could neither bind it or antecedent govern it. Therefore, generalized binding resulted in the ECP being reduced to a Principle A binding theory violation.

Within the Barriers framework, an antecedent in [NP,CP] c-commands a trace in subject position. The Minimality Condition blocks antecedent government, but is irrelevant for binding. (24)a is not a Principle A binding theory violation. Hence, within the Barriers framework, generalized binding does not eliminate the ECP.

14. Alternatively, the 'extracted' subject and pro might be generated in situ. The chief empirical distinction would be that base-generation should freely permit Subjacency violations. The evidence is not clear, but it appears that this structure in Somali is subject to the usual island conditions. For instance, NP cannot be extracted from a relative clause. (See (12) in section 4.)

15. German provides particularly clear examples, since WH-words are marked for Case in sentences like i. and ii.:

i. Wer hat den Hans gesehen?
Who has seen Hans?

ii. Wen hat der Hans gesehen?
Who(m) has Hans seen?

16. The external reviewer has observed, quite rightly, that my analysis evidently predicts that clause-bound moved NP should also be non-nominative, since Case is assigned at PF. A solution to this problem, consistent with the present analysis, is to make the (uncontroversial) assumption that PF has more than one level. Hence, focus and WH-movement occur at S-structure, Case assignment at PF, and clause-bound NP movement at PF'.

17. See Adams (1987, 17-19) for discussion and for citations within the generative literature.

18. Evidently, the direct relative is favoured, unless it results in ambiguity. See McCloskey (1979, 6-8).

19. It may not be entirely accurate to call aL and aN complementizers, since their distribution is quite unlike English that or French que, occuring as they do in simplex clauses. Moreover, they are themselves distinct from goN, the complementizer introducing ordinary sentential complements in Irish. What is important here is that aL and aN will be understood to be heads which occupy the position of C, head of CP.

20. I cannot explore here the many questions raised by the Irish data, e.g. how a resumptive pronoun in direct object position can be bound by an operator in [NP,CP] without violating Principle B.

21. Observe that if the synthetic form is available, either synthetic form (with pro) or analytic form (with lexical proroun) may appear. This is in marked contrast with base-generated pronouns, which must be pro, if the synthetic form is avaiable. (See M & H (p. 490), also quoted above in 5.1.)

In section 3.3, I proposed that the Avoid Pronoun Principle, repeated here as i., applies at D-structure.

i. Avoid pronoun if pro is available

The facts concerning base-generated pronouns in Irish fall out directly. If i. does not apply at S-structure, it will be possible for a resumptive pronoun created by subject extraction to be realized freely as pro or as lexical pronoun. Evidently, there is some cross-linguistic variation, since in Somali resumptive pronouns created by movement can only be realized as pro.

6. Conclusions

At the outset of this study, I proposed that there are basically two reasons for extending the theory to a new set of data, in this case a range of syntactic phenomena in Somali. The first is that the theory should provide a more explanatory account of the data, if indeed the theory makes correct generalizations about an innate human speech faculty. Secondly, in extending the theory, we inevitably encounter problems which challenge us to refine the theory itself. Of course, these are complementary processes, as the theory determines what questions we ask, and the data determines the range of possible answers.

It is appropriate at this point to evaluate what success has been achieved in meeting these twin objectives. Because this study entered new territory, it was necessary at times to deal with some peripheral issues, at the expense of the main line of investigation. Therefore, what follows is less a summary of the preceding sections, than an attempt to bring together those analyses which seem most secure and most interesting. It is convenient to retain a twin perspective and first summarize the conclusions regarding Somali syntax, then review the proposals of a more general and theoretical nature.

The co-occurence of extensive verb inflection paradigm, nominative Case assignment, and optional SCL suggests that Somali has a contrast between strong and weak subject-verb agreement, indicated here by AGR_S and AGR_W respectively. These co-occurence restrictions have been recognized to a greater or lesser extent by a range of authors, although attributing them to a unique AGR node is evidently original with this study. Given the known relationship between verb

inflection, nominative Case, and SCLs, the proposal is a natural one. However, the strongest support comes from the subsequent analyses, which have been constructed on this foundation. It is a crucial part of this analysis that SCL is generated <u>optionally</u> under AGR_S. It follows that those contexts where the SCL is either obligatory or excluded must be accounted for independently.

Given the above, it follows naturally that Somali is a null subject language in the sense that the subject position [NP,IP] may be occupied by pro, if certain conditions are met, in particular if the SCL is present. In section 3, I identified three contexts where SCL is necessary for giving content to pro: when pro is generated at D-structure, when pro is generated at D-structure but realized at PF as a lexical pronoun, and when clause-bound NP movement at PF leaves pro. In section 5, a fourth context was identified, when a subject is extracted from a sentential complement, leaving a pro which functions essentially as a resumptive pronoun. This part of the analysis accounted for when SCL is obligatory.

Finally, I identified a general prohibition against subject extraction from the domain of AGR_S, which is valid across null subject languages, including in particular Italian and Irish. As subject extraction from a simplex clause is only possible in Somali from the domain of AGR_W, and AGR_W cannot generate a SCL, it follows that SCL cannot occur in this context. This part of the analysis accounts for the contexts in which SCL cannot occur.

We see, therefore, that what had appeared in earlier work as complex and largely idiosyncratic phenomena can be made to follow from a distinction in Somali between AGR_S and AGR_W, requirements for the identification of pro, and a general prohibition against direct subject extraction from the domain of AGR_S.

In the course of this analysis, it was proposed that FP is generated under AGR (or I), along with an optional SCL if AGR=AGR_S. If the FP is a NP FP, the FP (and SCL, if present) move to C, head of CP, at S-structure, a species of head movement. This analysis seems reasonable, although it is not particularly interesting insofar as rather little can be made to follow from it as yet. Similarly, some proposals were made concerning Somali phrase structure, particularly the internal structure of VP, which are highly provisional and not very secure.

Now let us consider the theoretical consequences of this study. A striking feature of the Somali data is that it seems to force a distinction between AGR_S and AGR_W, to an extent unknown in other null subject languages. Italian lacks such a distinction, and the distinction in Irish is obscured by the fact that the synthetic form of the verb, corresponding to AGR_S, occurs only with pro. Consequently, AGR_W in Irish is unlikely to be perceived as an 'escape mechanism' for subject extraction, as it appears in Somali.

Subject extraction in null subject languages has been given varied interpretation. Rizzi (1982, c.4) attributes post-verbal subject extraction in Italian to that-t or ECP effects. However, the Somali data virtually forces the conclusion that AGR_S creates an opaque domain for direct subject extraction. There are at least two earlier studies which identify the problem as I have, namely Jaeggli (1984) and Roberge (1986c). Jaeggli (1984) makes incorrect predictions respecting subject extraction from sentential complement in Somali. As well, his Condition on Bound Pronominals is specific to a relatively narrow range of phenomena. Roberge (1986c) makes the correct empirical predictions, but introduces a condition on clitics that is essentially a stipulation.

The solution presented here, based upon the generalized binding of Aoun (1985), is an improvement over earlier accounts in that the restriction is attributed to a general principle, namely Principle B of the binding theory. Not only do we eliminate unnecessary conditions or principles, we also minimize the 'special status' of null subject languages.

The evidence from Italian is only partly consistent with these findings. Short subject extraction from a matrix clause is accounted for directly. However, subject extraction from sentential complement is problematic, since the extraction domain is structurally identical in Italian and Somali, yet subject extraction in Italian is from post-verbal position via free inversion. I assume this difference is due in part to independent factors which are peculiar to Italian. One possibility is that subjects in Italian are generated in post-verbal position.

The account presented here extends quite naturally to the Irish data, which is noteworthy, since the Irish facts are particularly complex, with a distinction between AGR_S and AGR_W, reflected in the 'synthetic' and 'analytic' forms of the verb, and three distinct complementizers, all of which interact to determine the extraction possibilities. Irish provides interesting, independent corroboration.

Finally, let us recall the crucial structure underlying subject extraction from sentential complement. (1) represents the structure created by subject extraction, before deletion of the intermediate trace:

(1) $\ldots[_{CP}\ t_i\ [_{C'}\ \underline{in}\ [_{IP}\ \ldots pro_i \ldots AGR_S \ldots$

A subject cannot be extracted from the domain of AGR_W in a structure like (1), because AGR_W does not properly govern a trace and the intermediate trace cannot antecedent govern the extraction site, by the Minimality Condition. Therefore, subject extraction leaves a pro, which is licensed by AGR_S and

must be given content, i.e. requiring presence of SCL. The intermediate trace does not antecedent govern pro, but it does bind it. Therefore, the intermediate trace must delete before binding theory.

This analysis relies upons the Barriers framework, which crucially permits an intermediate trace to bind, but not govern, the subject position, if a suitable head in C invokes the Minimality Condition. This makes precisely the correct prediction in the Somali case of focus extraction of a subject NP from a sentential complement. Since the intermediate trace is deleted before binding theory, the extracted subject is not assigned nominative Case.

At the same time, subjects may be extracted over a non-governing head, e.g. the NP FP baa or ayaa (alternatively, b- or ay-) in Somali, or the complementizer aL in Irish. These distinctions were not readily available in pre-Barriers frameworks.

I have surveyed a range of phenomena in Somali syntax and have attempted to account for them within the Barriers version of government binding theory. To the extent that features of Somali syntax have been correctly attributed to the human speech faculty, the present account achieves an improved level of explanatory adequacy. Somali has also provided strong evidence for some modifications to the theory. Only future study will determine whether these proposals can be maintained. Whatever the eventual judgment, it is to be hoped that the analysis of Somali syntax presented here will in some way contribute to the larger enterprise.

7. Bibliography

Abraham, R.C. (1967) English-Somali Dictionary, University of London Press, London.

Adams, M. (1987) "From Old French to the Theory of Pro-Drop," Natural Language and Linguistic Theory 5, 1-32.

Andrzejewski, B.W. (1961) "Notes on the Substantive Pronoun in Somali," African Language Studies 2, 80-99.

_____ (1964) The Declensions of Somali Nouns, School of Oriental and African Studies, University of London, London.

_____ (1968) "Inflectional Characteristics of the So-called 'Weak Verbs' in Somali," African Language Studies 9, 1-51.

_____ (1969) "Some Observations on Hybrid Verbs in Somali," African Language Studies 10, 47-69.

_____ (1975) "The Role of Indicator Particles in Somali," Afroasiatic Linguistics 1, 1-69.

_____ (1984) "The Role of Accentual Patterns in Subject/Object Differentiation in Somali and Its Parallels in Paranilotic Languages," in J. Byron, ed. Current Progress in Afro-Asiatic Linguistics, Benjamins, Amsterdam, 11-15.

Antinucci, F. (1980) "The Syntax of Indicator Particles in Somali, Part Two: The Construction of Interrogative and Negative Interrogative Clauses," Studies in African Linguistics 11, 1-37.

_____ and A. Puglielli (1980) "The Syntax of Indicator Particles in Somali: Relative Clause Construction," Afroasiatic Linguistics 7, 85-102.

_____ and A. Puglielli (1984) "Relative Clause Construction in Somali: A Comparison between the Northern and Coastal Dialects," in J. Bynon, ed. Current Progress in Afro-Asiatic Linguistics, Benjamins, Amsterdam, Holland, 17-31.

Aoun, J. (1985) A Grammar of Anaphora, Linguistic Inquiry Monograph 11, MIT Press, Cambridge, MA.

Bell, C.R.V. (1953) The Somali Language, Longmans, London. Reprinted by Gregg, 1968.

Belletti, A. and L. Rizzi (1981) "The Syntax of 'ne': Some Theoretical Implications," Linguistic Review 1, 117-154.

Borer, H. (1981) "On Extraction from Clitic Doubled Constructions," NELS 11, 22-37.

_____ (1984) Parametric Syntax: Case Studies in Semitic and Romance Languages, Studies in Generative Grammar 13, Foris Publications, Dordrecht, Holland.

Brandi, L. and P. Cordin (1981) "On Clitics and Inflection in Some Italian Dialects," ms., Scuola Normale Superiore, Pisa.

_____ (1989) "Two Italian Dialects and the Null Subject Parameter," in Jaeggli and Safir, eds. The Null Subject Parameter, Kluwer, Dordrecht, 111-142.

Brody, M. (1984) "On Contextual Definitions and the Role of Chains," Linguistic Inquiry 15, 355-380.

Carroll, S. (1982) "Redoublement et dislocation en francais populaire," in C. Lefebvre, ed. La Syntax comparee du francais standard et populaire: approches formelles et fonctionelles, Office de la langue francaise, Quebec, t. 1, 291-357.

Chao, W. (1981) "Pro-Drop Languages and Non-Obligatory Control," ms., University of Massachusetts, Amherst.

Chomsky, N. (1965) Aspects of the Theory of Syntax, MIT Press, Cambridge, MA.

_____ (1976) "Conditions on Rules of Grammar," Linguistic Analysis 2, 303-351.

_____ (1980) "On Binding," Linguistic Inquiry 11, 1-46.

_____ (1981) Lectures on Government and Binding, Foris Publications, Dordrecht, Holland.

_____ (1982) Some Concepts and Consequences of the Theory of Government and Binding, Linguistic Inquiry Monograph 6, MIT Press, Cambridge, MA.

_____ (1986a) Barriers, Linguistic Inquiry Monograph 13, MIT Press, Cambridge, MA.

_____ (1986b) Knowledge of Language: Its Nature, Origin and Use, Praeger, New York.

_____ (1988) "Some Notes on Economy of Derivation and Representation," ms., MIT, Cambridge, MA.

Doron, E. (1986) "The Pronominal 'Copula' as Agreement Clitic," in H. Borer, ed. The Syntax of Pronominal Clitics, Syntax and Semantics 19, Academic Press, Orlando, FL, 313-332.

El-Solami-Mewis, Catherine (1980) "Zur Rolle der Satzpartikeln beim Ausdruck des Logischem Pradikats im Affirmativem Somalisatz," Zeitschrift fur Phonetik, Sprachwissenshaft und Kommunicationsforschung 33, 105-127.

_____ (1987) Lehrbuch des Somali, Verlag Enzyklopadie, Leipzig.

Everett, D.L. (1987) "Pirahã Clitic Doubling," Natural Language and Linguistic Theory 5, 245-276.

George, L.M. and J. Kornfilt (1981) "Finiteness and Boundedness in Turkish," in F. Heny, ed. Binding and Filtering, MIT Press, Cambridge, MA, 105-127.

Hetzron, R. (1965) "The Particle baa in Northern Somali," Journal of African Languages 4, 118-130.

Horvath, J. (1986) Focus in the Theory of Grammar and the Syntax of Hungarian, Foris Publications, Dordrecht, Holland.

Hyman, L. (1981) "Tonal Accent in Somali," Studies in African Linguistics 12, 169-203.

_____ and J.R. Watters (1984) "Auxiliary Focus," Studies in African Linguistics 15, 233-273.

Jaeggli, O. (1982) Topics in Romance Syntax, Foris Publications, Dordrecht, Holland.

_____ (1984) "Subject Extraction and the Null Subject Parameter," NELS 14, 132-153.

_____ (1986) "Three Issues in the Theory of Clitics: Case, Doubled NPs, and Extraction," in H. Borer, ed. The Syntax of Pronominal Clitics, Syntax and Semantics 19, Academic Press, Orlando, FL, 15-42.

Jaeggli, O. and K. Safir (1989) "The Null Subject Parameter and Parametric Theory," in Jaeggli and Safir, eds. The Null Subject Parameter, Kluwer, Dordrecht, 1-44.

Kayne, R. (1975) French Syntax: The Transformational Cycle, MIT Press, Cambridge, MA.

_____ (1989) "Null Subjects and Clitic Climbing," in Jaeggli and Safir, eds. The Null Subject Parameter, Kluwer, Dordrecht, 239-261.

Kenstowicz, M. (1989) "The Null-Subject Parameter in Modern Arabic Dialects," in Jaeggli and Safir, eds. The Null Subject Parameter, Kluwer, Dordrecht, 263-275.

Koopman, H. (1984) The Syntax of Verbs: From Verb Movement Rules in the Kru Languages to Universal Grammar, Foris Publications, Dordrecht, Holland.

_____ and D. Sportiche (1982) "Variables and the Bijection Principle," Linguistic Review 2, 139-160.

Lasnik, H. and M. Saito (1984) "On the Nature of Proper Government," Linguistic Inquiry 15, 235-289.

Lecarme, J. (1984) "On Somali Complement Constructions," in T. Labahn, ed. Proceedings of the Second International Congress of Somali Studies, University of Hamburg, August 1-6, 1983: Vol. I, Linguistics and Literature, Helmut Buske Verlag, Hamburg, 37-54.

Livnat, M.A. (1984) Focus Constructions in Somali, PhD. Dissertation, University of Illinois at Urbana-Champaign.

Luling, V. (1987) Somali-English Dictionary, Dunwoody Press, Wheaton, MD.

McCloskey, J. (1979) Transformational Syntax and Model Theoretic Semantics : A Case Study in Modern Irish, D. Reidel, Dordrecht.

_____ and K. Hale (1984) "On the Syntax of Person-Number Inflection in Modern Irish," Natural Language and Linguistic Theory 1, 487-533.

May, R. (1977) The Grammar of Quantification, PhD. Dissertation, MIT.

Mohammed, M.A. (1985) "Subject Extraction in Arabic," Proceedings of the West Coast Conference on Formal Linguistics 4, 220-227.

Murrell, M. and V. Ştefănescu-Drăgăneşti (1970) Romanian (Teach Yourself Books), Hodder and Stoughton, London.

Owens, J. (1985) A Grammar of Harar Oromo (Northeastern Ethiopia), Including a Text and a Glossary, Cushitic Language Studies 4, Helmut Buske Verlag, Hamburg.

Perlmutter, D. (1971) Deep and Surface Structure Constraints in Syntax, Holt, Rinehart and Winston, New York.

Pollock, J.-Y. (1989) "Verb Movement, Universal Grammar, and the Structure of IP," Linguistic Inquiry 20, 365-424.

Reinhart, T. (1983) Anaphora and Semantic Interpretation, University of Chicago Press, Chicago, IL.

Rizzi, L. (1982) Issues in Italian Syntax, Foris Publications, Dordrecht, Holland.

_____ (1986) "Null Objects in Italian and the Theory of pro," Linguistic Inquiry 17, 501-557.

Roberge, Y. (1986a) "Subject Doubling, Free Inversion, and Null Argument Languages," Canadian Journal of Linguistics 31, 55-79.

_____ (1986b) "On Doubling and Null Argument Languages," NELS 16, 388-402.

_____ (1986c) The Syntactic Recoverability of Null Arguments, PhD. Dissertation, University of British Columbia.

Rochemont, M.S. (1986) Focus in Generative Grammar, Studies in Generative Linguistic Analysis 4, Benjamins, Amsterdam.

Saeed, J.I. (1984) The Syntax of Focus & Topic in Somali, Cushitic Language Studies 3, Helmut Buske Verlag, Hamburg.

_____ (1987) Somali Reference Grammar, Dunwoody Press, Wheaton, MD.

Serzisko, F. (1984) Der Ausdruck der Possessivität im Somali, Continuum: Schriften Reihe zur Linguistik 1, Gunter Narr Verlag, Tubingen.

Sportiche, D. (1983) Structural Invariance and Symmetry in Syntax, PhD. Dissertation, MIT.

Stowell, T. (1981) Origins of Phrase Structure, PhD. Dissertation, MIT.

Travis, L. (1984) Parameters and Effects of Word Order Variation, PhD. Dissertation, MIT.

Zwicky, A.M. (1985) "Clitics and Particles," Language 61, 283-305.

_____ and G.K. Pullum (1983a) "Phonology in Syntax: The Somali Optional Agreement Rule," Natural Language and Linguistic Theory 1, 385-402.

_____ and G.K. Pullum (1983b) "Cliticization vs. Inflection: English n't," Language 59, 502-513.