

Somali as a Subject *Pro-drop* Language:
Similarities Between Spanish and Somali in the Syntactic Structure of the Subject Position

A PLAN B PAPER
SUBMITTED TO THE FACULTY OF
THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA
BY

Tara Leigh Gibbs

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTERS OF ARTS IN ENGLISH AS A SECOND LANGUAGE

July 2007

Plan B paper ready for defense:

Project Supervisor

Date

Abstract

Abstract

This paper suggests that the reason Somali and Spanish speaking students both frequently delete subject pronouns in English is because they are both *pro-drop* languages. It reanalyzes the so-called ‘subject clitic pronoun’ on sentential markers as an agreement marker which licenses *pro-drop*. As a by-product of the analysis, it also suggests that CP (formerly known as the complementizer phrase) is a head-initial phrase, while IP (formerly known as the inflectional phrase) is a head-final phrase.

Introduction

Have you ever looked at the writing of Advanced Beginning/Low Intermediate Somali ESL students and been surprised that they leave off the subjects in sentences the same way Spanish speaking students do? Is this a general trait of language acquisition, or is there something similar about Spanish (an Indo-European, Romance language), and Somali (an Afro-Asiatic, Cushitic language)? Russian speaking students don’t seem to do this, nor do French speaking students or Hmong students. What then is going on here? What could Somali speakers and Spanish speakers have in common that causes them to do similar things when constructing English sentences?

As a teacher I frequently struggle with what grammatical errors to correct and what not to correct. As a general rule, I find language transfer errors to be more worthy of correction than errors which are part of a universal language acquisition process which just needs time to develop, so knowing whether Somali students are leaving off subjects because it is a universal learning process or because it is a transfer error affects my choices in teaching.

This paper examines the structure of Somali and suggests that Somali, like Spanish, is a *pro-drop* language, illustrated below, in which pronominal subjects are often realized through agreement, rather than through an overt, independent pronoun. Thus, Spanish speakers usually say “Hablo español” rather than “Yo hablo español for “I speak Spanish”. This paper reanalyzes the marking on the sentential markers in Somali (such as, *baa*, *ayaa*, *waa*, *waxaa*), as agreement markers. It suggests that *pro-drop* is the common structural source for transfer errors in the writing of both Spanish and Somali ESL students who regularly omit the subject in their writing. As a by-product of this analysis, this paper also looks at the head ordering in CP (formerly the complementizer phrase) and IP (formerly the inflectional phrase).

Previous descriptions of Somali were developed prior to the instantiation of a solid understanding of the theory of *pro-drop* within the field of linguistics (although many were published a decade or more later), and as a result they have tried to analyze the subject agreement marker on the various sentential markers as the subject pronoun (Saeed (1999), Mansur and Puglielli (1999), Lamberti (1988), and Andrzejewski (1984, 1979, 1975a, 1975b, 1974, 1964, 1961)). The theory of *pro-drop*, which is described later in the paper, was developed in the mid-80’s and remained controversial into the late 80’s and early 90’s (Jaeggli & Safir 1989). Textbooks of this period felt compelled to describe the controversy (Haegemon 1991), where as current textbooks describe this phenomena uniformly as *pro-drop*, regardless of their theory of instantiation (Ouhalla 1999:p 320; Van Vallin 2001:p101).

The ability of *pro-drop* to conceptually unify accounts of language variation cross-linguistically and to eliminate the need for multiple, language specific accounts of

the phenomena, as well as its identification in many languages and its cross-linguistic explanatory power of previously unexplained phenomena, has caused it to rise to the level of a cross-linguistic construction which every theory of grammar must be able to account for, whether they instantiate it as *pro*, or in another structural way. Thus, Lexical Functional Grammar (LFG), Head-driven Phrase Structure Grammar (HPSG), and all theories of grammar with which I am familiar, offer accounts of this phenomena which is consistent across languages and consistent with the mechanics of their theory (Pre-session lecture on LFG at the LSA Institute presented by Mary Dalrymple, July 2, 2007; Pre-session lecture on HPSG at the LSA Institute presented by Ivan Sag, July 3, 2007).ⁱ

Work on Somali has not taken this phenomena into account, principally because the work on Somali flourished in the mid-80's, with its authors trained prior to that time, and its earliest authors imposed the requirements of their own native languages (a consistent, overt instantiation of the subject) on the system. Since the beginning of the civil war in Somalia in 1992 little new work has been undertaken on the linguistic structures of the Somali language, although older works, such as Mansur and Puglielli's work, which was to have been published in the early 90's by the Somali government, did eventually make its way to press in Europe in the late 90's.

The first section of this paper will describe the study methodology, the linguistic relationship of Somali to other languages, and dialect divisions of Somali. The second section of this paper will review linguistic theories pertaining to the analyses in this paper and the literature on Somali. It will discuss the theta-criterion, assignment of theta-roles, case theory, movement theory, and the processes for assigning agreement. It will look at the evidence for an non-overt pronoun (*pro*), and it will also describe the characteristics

and licensing of *pro*. Finally, it will look at prior analyses of the subject-clitic pronoun system in Somali, analyses of the declarative marker, topic marker, and interrogative marker by Saeed, and Saeed's analysis of the sentence structure. The third section of this paper will argue for reinterpretation of the subject-clitic pronoun system as agreement marking, and the status of Somali as a *pro-drop* language based on the optionality of the free subject pronoun. It will also suggest that IP (which at one time meant inflectional phrase, but no longer does) is a head-final phrase based on the positioning of the tense and aspect markers, while CP (which at one time meant complementizer phrase, but no longer does), is head initial, resulting in a split-head parameter for Somali, an analysis that has been offered for some other languages with SOV order and left complementizers, such as German. Finally, this paper will look at the two morphemes representing subject agreement and use them to motivate movement of the subject from spec-VP to spec-IP to spec-CP.

Background: Data Selection, Language and Dialect Divisions, and Description

Data Selection

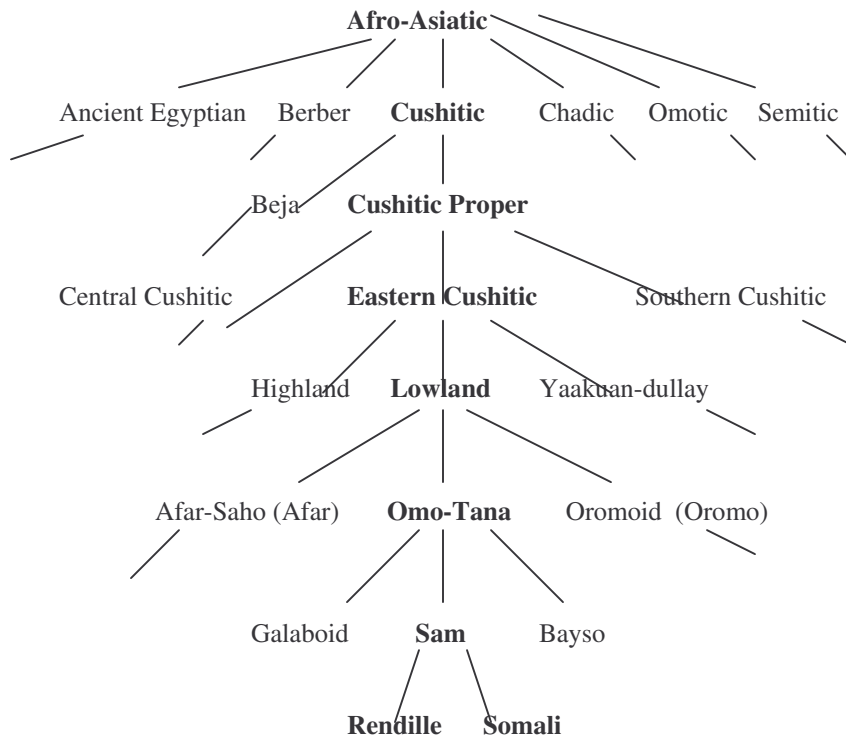
Although there is considerable variation across the Northern Dialect of Somali with regard to data relevant to analyzing the object system in Somali, with regard to the data needed to analyze the subject position there is great uniformity. For this reason, none of the data sets representing Standard Somali were excluded from the analysis in this paper as representing contradictory judgments or different underlying structures in different dialects. Most of the data in this analysis replicates paradigms that can be found in any introductory textbook about Somali. The examples cited in this study were all

verified via grammaticality judgment questionnaires and grammaticality judgment elicitations with multiple speakers from a variety of regions including Hargessa, Basaso, Mogadishu, and Northern Kenya (Mandera).

Language Divisions

Somali is a member of the Lowland, East branch of the Cushitic language family, as seen in Figure 1 below. The closest related language is Rendille, which, like Somali, is an Omo-tana, Sam language (Ruhlen, 1991). Other well-known Lowland languages are Oromo, spoken in Ethiopia, and Afar, spoken in Ethiopia, Djibouti and parts of Somalia. The language tree for Somali, pictured below, is based on the work of Ruhlen (1991) with a slight change to the relationship of Rendille and Somali based on Saeed (1999). There are approximately 7.75 million speakers of Somali in Somalia, 3.3 million

Figure 1: Language Tree for Somali (Based on Ruhlen 1991 and Saeed 1999)



in Ethiopia, and another 0.75 million living in other countries, including Djibouti, Kenya, Norway, the United States, Canada, and the United Arab Emirates. Work on the classification of Somali dialects can be found in Ehret and Mohamed N. Ali (1984), Lamberti (1986, 1988), and Diriye (2001). Appendix C includes a translation of the introduction to Lamberti's 1988 book *Die Nordsomali-Dialekte* which gives a comprehensive overview of work on the classification of all dialects of Somali through the mid-80's.

Dialect Division

Saeed (1999) divides Somali into three main dialects, Northern, Benadir and Maay, and several smaller dialects, Jiddu, Tunni, Brava, Karre, Ashraf, and Merka. Many Somalis in Minnesota refer to these three dialects instead as Common/Standard Somali, Coastal (Benadir) Somali, and Maay, respectively.

The Northern dialect has spread via migration to the northern, western, and southern areas of Somalia, parts of Ethiopia and Kenya. Saeed claims that despite its large geographical coverage, there is "relative homogeneity of Standard [Northern] Somali" (P. 5). When evaluated by the standard of *mutual intelligibility*, this may be true; however, I have found considerable variation in grammaticality judgments by speakers claiming to speak Standard Somali regarding structures not analyzed in this paper, as well as regarding a variety of structures of interest when evaluating the quality of translations, a topic which will not be discussed in this paper. None of these structures seem to pose any intelligibility problems; however, they spark huge debates regarding "purity" and "correctness", which suggests that a socio-linguistically defined standard

variety of the Somali language does not yet exist among the dialects of the Somali language, causing the term *Standard Somali* to be something of a misnomer from a socio-linguistic point of view.

Language Description

Relatively little linguistic research has been done on the Somali language. The most comprehensive grammars have been written by Saeed (1999), Lamberti (1988), and Mansur and Puglielli (1999). Mansur and Puglielli's work (1999) is based on an early version of transformational grammar which does not utilize x-bar theory or the theory of adjunct and complement structure. Saeed's grammars of Somali (1999, 1986) do not utilize a theory of transformational grammar or the concept of a verb phrase. Instead, he utilizes the descriptive concepts of a verbal core and satellites working in constellation in a verbal group, reminiscent of Dik's (1989) theory of Functional Grammar. Likewise, the grammar book written by Zorc and Issa (1990) and the work of Andrzejewski do not utilize theories of generative grammar.

Literature Review

Theoretical Background

Theta Roles and the Theta Criterion.

Current theories of grammar recognize that semantic, phonological, and syntactic properties interact with each other in constructing a grammatical utterance. One of the interfaces between the semantic properties of words and the syntactic rules of the language is currently known as Theta-theory. This theory recognizes that verbs specify

particular types of noun phrases, prepositional phrases, and sentential clauses (collectively, complements) that can be associated with them and assign them relationships to the verb. For example, in the sentence “Alex hit Darius” ‘Alex’ is the agent of the hitting and ‘Darius’ is the recipient of the hitting. These roles, called *agent* and *patient* respectively, are referred to as thematic roles, or theta-roles for short. The theories which label the relationships of different constituents to the verb for each verb in a language are based in semantic theory, not syntactic theory; however, syntax recognizes that without a theta-role, a constituent phrase has no function in the sentence. The verb *hit* assigns two theta-roles: Agent and Patient, which the syntax in English places before and after the verb respectively. Therefore, “Alex hit Darius tornado” makes no sense because the relationship of ‘tornado’ to the sentence is unknown and thus uninterpretable. In other words, in this sentence, ‘tornado’ has no theta-role assigned to it. The theta criterion consists of two rules developed within the theory which are stated in (1).

(1) Theta-Criterion

- i) Each argument must be assigned a theta-role.
- ii) Each theta-role must be assigned to an argument.

Argument refers to a semantic property. It refers to the participants involved in the event denoted by the predicate (Ouhalla 1999, p. 148). In a sentence like “I like chocolate”, the arguments of ‘like’ are ‘I’ and ‘chocolate’. The implication of the Theta-Criterion is that a sentence like “Alex hit Darius tornado” is ungrammatical, because the argument ‘tornado’ has not been assigned a theta-role. Likewise, a sentence like “hit Darius” is also ungrammatical, because the theta-role of ‘agent’ has not been assigned to an argument. Instead, a structure must be selected which does not have an agent theta role

to assign, such as the passive construction “Darius was hit.” Similarly, the Theta-Criterion rules out sentences like “Alex hit him Darius”, because the theta-role of ‘patient’ can only be assigned once to one argument and cannot be assigned twice to both ‘him’ and ‘Darius’. Another example of a sentence ruled out in the same way is “Alex hit him him”.

Case Assignment.

While semantics defines the set of theta-roles for each verb in the language, syntax defines the set of cases for each verb. Thus, there are verbs like “jump”, which are intransitive, meaning the only case associated with it is nominative. There are also verbs like “lock” which are transitive and have both nominative and accusative arguments associated with it. “I gave John the book” is an example of a di-transitive verb which has nominative, accusative, and dative arguments associated with it. Case theory, which is described in detail in Ouhalla (1999), assigns case as follows:

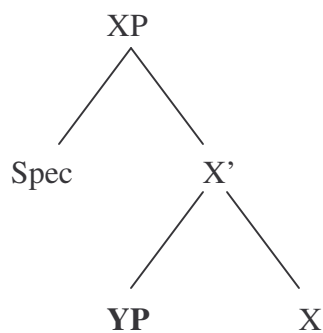
(2) Case Assignment

- (A) Accusative (dative etc) case: is assigned by the verb to its complement.
- (B) Nominative case: is assigned by INFL to the noun phrase in its specifier position

INFL (I), which is the head of IP (formerly called the inflectional phrase, but now viewed as the locus of finite vs. non-finite distinction, not inflection which is just one way finiteness can be realized) is equivalent to S in earlier versions of transformational grammar; however, unlike S, IP is assumed to have the same x-bar structure as other phrases, such as NP or VP. X-bar theory posits that any phrase has a head, which is

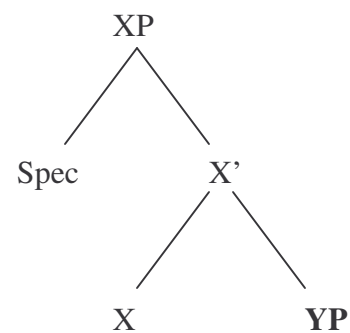
generically called X, because it could refer to N (a noun), V (a verb), P (a preposition) and so forth. X projects up to XP, the phrase level. For example, N projects up to NP (the noun phrase), and V projects up to VP (the verb phrase). Between X and XP is an intermediate level, X', pronounced X-bar. X' is the level at which complements are adjoined to the phrase. Complements (YP) can branch to the left, as seen in Figure 2, or to the right, as seen in Figure 3. When complements branch to the left, the phrase is called head-final. When they branch to the right, the phrase is called head-initial. Additionally, Figures 2 and 3 show a specifier position. The specifier position is not well understood. It simultaneously displays characteristics of adjunction, while appearing bound to the phrase fundamentally, participating in control and government of a phrase. The specifier position apparently only branches to the left, because, apparently, there are no examples of languages in which it branches to the right; however, no theoretical reason for this has yet been found. Figure 2 is the generic footprint for a head-final phrase and Figure 3 is the generic stamp for a head-initial phrase.

Figure 2



Head-final phrase

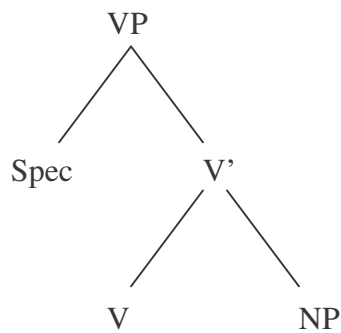
Figure 3



Head-initial phrase

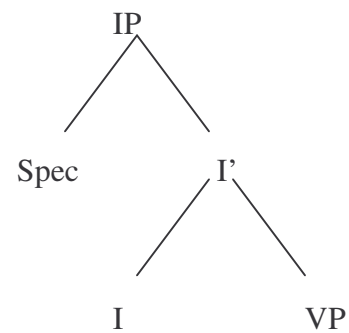
In the verb phrase pictured in Figure 4 below, you see the verb phrase (VP), V' (pronounced V-bar), and the head V. The object noun phrase branches off from V' to the right, so that the V-head is on the left, making this a head-initial verb phrase. Figure 5 is an example of a head-final IP projection of the head I. Common heads for IP include tense, the third person singular -s, aspect, and mood, such as modal verbs like 'can'. Like other phrases, IP has a specifier position, and is of the form XP-X'-X.

Figure 4



Head-initial Verb Phrase

Figure 5



Head-final Verb Phrase

For a decade or two, English was described as having its subject originate in the specifier position of IP, and its object originate in the complement of VP. This, however, was unsatisfying, since it seemed that the subject and object were both closely related to the verb and should therefore both originate in the verb phrase where they received theta-roles. This eventually led to the suggestion that the subject originates in the specifier of the verb [spec-VP], but then moves to [spec-IP] in order to get a case assignment. Case is assigned via a condition called:

(3) Spec-Head Agreement

A head (X) and its specifier (Spec-XP) must agree in relevant features.

Thus, the subject is attracted to the specifier of IP ([spec, IP]) and moves there, because it lacks case and needs to get it. The [spec, IP] position can satisfy this need (see 2B above).

Movement

Besides moving to get case, there are various other things that can motivate movement in Movement Theory. CP (formerly called the complementizer phrase, but now viewed as the locus of information-structure motivated adjunction, as well as subordination and discourse moods such as the interrogative, conditional, and so forth—does it seem like CP has become a dumping ground for all context-dependent grammatical items which needs to be refined into several layers) also drives movement. For example, a feature [Q] is posited to be a feature of the head of CP in interrogative sentences in all languages. CP is equivalent to S' in earlier versions of transformational grammar; however, unlike S', CP has the same x-bar structure as other phrases, such as NP or VP. That is, CP is a projection of the head C. 'For' and 'that' in English, and 'ka', the question marker in Japanese, are common heads of CP. Like other phrases CP has a specifier position, and is of the form XP-X'-X. To restate, [Q] is an example of a Phi-feature, which is a feature that must be matched with an element in the sentence.

In English, the Phi-feature [Q], which is possessed by C in interrogative sentences, has the effect of attracting the WH-phrase to [Spec-CP] because [Q] and the

WH-phrases have like features which must be in a spec-head relationship. The implication of this is that in a sentence like

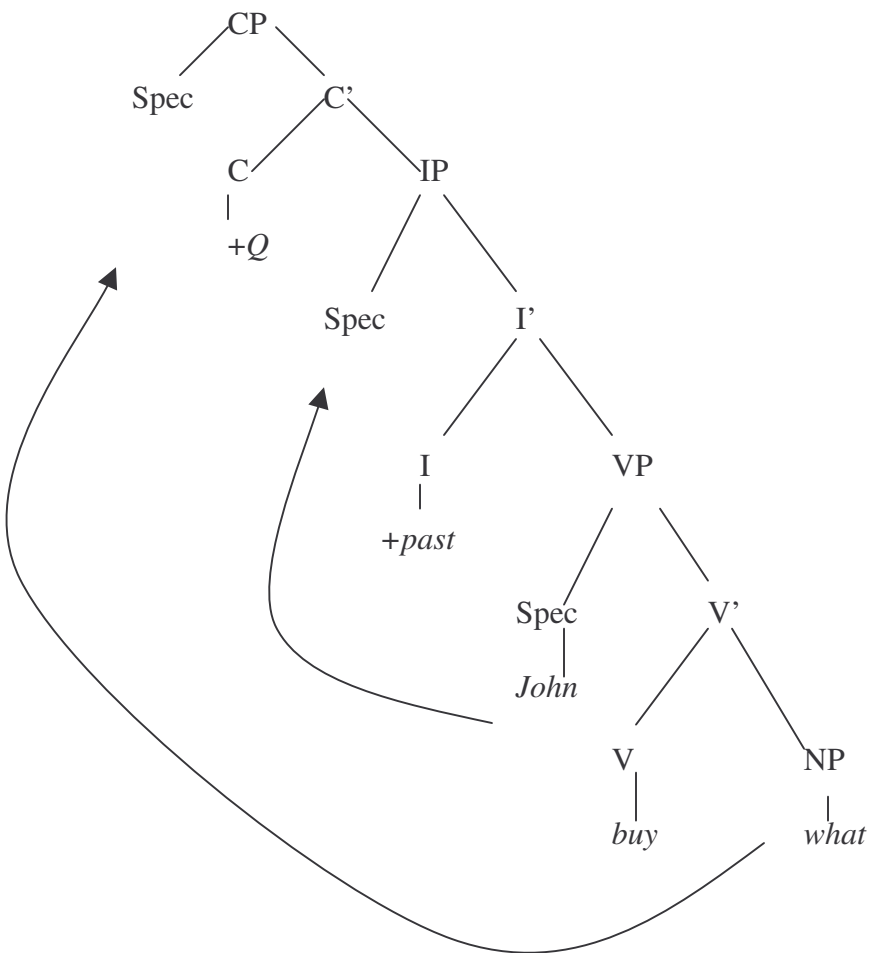
(4) “What did John buy?”

the original structure (called Deep Structure) of this sentence is

(5) Q +past +John buy what

Restated formally as:

(6) $[_{CP} Q [_{IP} +past [_{VP} [John [_{V'} [buy][what].$



In this diagram, the word “what” is attracted to the [Q]-feature in CP and moves to the specifier of CP in order to achieve spec-head agreement between the Q-morpheme ‘what’ and the Q-feature in the head of C, leaving behind something referred to as a trace. To restate, features which must be matched, such as the Q-feature in the head of C and the Q-morpheme ‘what’ are often called Phi-features, and it is the existence of these Phi-features which are said to motivate the Wh-movement in this sentence, just as the movement of ‘John’ from the specifier of the verb to the specifier of IP is motivated by the need for ‘John’ to have case assigned to it. As described above, the specifier position of the verb is understood to receive theta marking from the verb, but not case marking. Nominal case marking can only be assigned by IP. Thus, the subject must move out of the verb phrase and into the specifier of IP in order to get case via a process of Spec-Head agreement.

Agreement

Agreement features, such as –s in “John buys...” are also assigned via Spec-head agreement, so the head, I, must agree in phi-features with its specifier (in this case ‘John’). Thus, since ‘John’ is singular, masculine, and third person, these phi-features must also be the phi-features of the head of I, generating the –s marker rather than a null marker.

Pro

In the early 1980’s, research in Chomsky’s generative grammar theory, known as Government and Binding theory, an evolution of a theory known as Principles and

Parameters, developed evidence for three null categories which had the feature matrices illustrated in Figure 2 below. Null categories are categories which have no phonological realization, but which act as operators in the syntax the same way overt categories do.

Figure 6: Feature matrices of three null categories

DP-trace	Variable-trace	PRO
[+anaphor]	[-anaphor]	[+anaphor]
[-pronoun]	[-pronoun]	[+pronoun]

The DP-trace is illustrated by the gap left in a relative clause construction. An example of this in English is “The boy that I saw was very tall”. This is posited to be of the form “[The boy_i [that I saw t_i] was tall]” where the *the boy* and *t* represent the same entity, but the object in the relative clause is ‘missing’. The variable-trace is illustrated by the sentence “Who did you see?” where the object of see is ‘missing’. Underlyingly this is posited to be of the form [Who_i [did you see t_i?] PRO is the pronoun that is found in the subject position of untensed clauses such as “I was glad to see you.” Underlyingly this is posited to be of the form [I_i was glad [PRO_i to see you]].

Missing from this paradigm was a null category with the properties [-anaphor] and [+pronoun] which had the properties of a pronoun, but which was not pronounced in the output of the sentence and had no anaphoric reference in the sentence, for example a sentence like ‘ate apples’ meaning ‘(I) ate apples’. Rizzi (1986a, 1986b) suggests that this category exists in languages such as Spanish and Italian where the subject morpheme is not obligatory, but where the sentences are judged as being grammatical despite the fact that there is no apparent noun phrase to take the nominative case and agent theta-roles being assigned. For example in (7) below:

- (7) Ha telefonato.
Has phoned.
“He/she has phoned.”

Rizzi claims that this structure underlyingly is

- (8) [_{IP} *pro* _I [_{VP} ha telefonato].

Not all languages can replace their subject with a *pro* null category. English, for example, cannot replace its overt noun phrase with a *pro* form. ‘has phoned’, meaning ‘He has phoned’ is not grammatical. The conditions that need to apply in order for a construction (for example, the construction *pro-drop*) to be allowed are called the licensing conditions. According to Rizzi (1986a), in order to license *pro*, the following condition must be met:

- (9) Condition on the licensing of *pro*
Pro is licensed by an overt (rich) Agreement category co-indexed with it.

This condition explains why English, an extremely agreement-poor language, cannot generally license *pro*.ⁱⁱ Another principle of syntax which might be utilized to license *pro* could be the principle of recoverability, which essentially says the same thing in more general termsⁱⁱⁱ.

Languages which can license *pro* as a general rule^{iv} are called *pro-drop* languages. *Pro-drop* languages seem to have a number of other traits that tend to co-occur (Taraldsen 1981, Rizzi 1986a). One of these traits is the trait of *free inversion*. Free inversion is a phenomena in which the “overt subject of a finite clause can appear freely in the clause final position following the verb and other VP-material.” (Ouhalla

1999, P. 314) Thus, you can get the sentence seen in (10) which places the subject NP at the end of the sentence.

- (10) Ha telefonato Gianni
Has telephoned Gianni
'Gianni has phoned.'

Clitics and Agreement

Phonological clitics refer to semantically independent morphemes, which are not affixes, that adjoin themselves to another morpheme forming a single phonological word. Clitics exist in most languages. In English, the determiners “the” and “a”, and the preposition “to”, for example, often cliticize to other words. Thus, in the sentence “I want to go”, ‘want’ and ‘to’ often cliticize and become a single phonological word, written as ‘wanna’ in cartoons, in order to show their transformation into one word. The term “syntactic clitics” refers to movement of a particular type of morpheme to a designated position in the sentence.

Ouhalla (1999) describes three types of clitic languages: variable clitic languages, clitic second languages, and fixed position clitic languages. Variable clitic languages can put their clitic pronouns in various positions, including at the beginning of sentences and before and after the verb (French is an example). Clitic second languages, like Portuguese, must always have some other element at the beginning of the sentence. The clitics often seem to occur before the verb, but not always. Fixed clitic languages have clitics in a position which is fixed relative to a particular category. Examples of fixed clitic languages are Arabic and Hebrew. Roberts and Shlonsky (1995) argued that these

fixed clitics are actually agreement markers co-referenced with a particular category of the language.

The Verb in Somali

Somali is an agreement rich language, showing verbal inflections which agree for person and number with the subject. Traditional grammars of Somali base the conjugation of verbs on the base form of the verb which happens to correspond to the imperative form in Somali (Orwin 1995, Zorc and Issa 1990). Examples of verbs in Somali in their base are *tag* (go), and *cun* (eat).

Somali verbs inflect to show tense and agree for person with the subject, as can be seen in examples 11 and 12 which illustrate the first and second person singular forms of *cun* (eat) and *tag* (go) (Orwin 1995, Zorc and Issa 1990).

11a Aniga waan tag-**aa**
 I DM-1st per sg Go-1st per sg pres
 'I go'

11b Aniga waan cun-**aa**
 I DM-1st per sg Eat-1st per sg pres
 'I eat'

12a Adiga waad tag-**taa.**
 You DM-2nd per sg Go-2nd per sg pres
 'You go'

12b Adiga waad cun-**taa.**
 You DM-2nd per sg Eat-2nd per sg pres
 'You eat'

(DM stands for Declarative marker)

As you can see, the ending of the verb changes to reflect person and tense. If you contrast these with 15 and 16 below, you will also see that they reflect number. In examples 13 and 14 you can see that in the third person singular, the verb also agrees for gender, as well as person, number and tense.

13a Isaga wuu tag-**aa**
 He DM-3p sg masc Go-3p sg masc pres
 'He goes'

13b Isaga wuu cun-**aa**
 He DM-3p sg masc Eat-3p sg masc pres
 'He eats'

14a Iyada way	tag- taa	14b Iyada way	cun- taa
She DM-3p sg fem	Go-3p sg fem pres	She DM-3p-sg fem	Eat-3p sg fem pres
'She goes'		'She eats'	

In examples 15-18 you can see that the paradigm also shows number agreement. 15 and 16 are inclusive and exclusive 'we', a contrast which seems to be disappearing from the language and which may not be familiar to all Somali speakers.

15a Innagu waynu	tag- naa .	15b Innagu waynu	cun- naa .
We(incl) DM-1 st per incl pl	Go-1 st p pl pres	We(incl) DM-1 st per incl pl	Eat-1 st per pl pres
'We (incl) go'		'We (incl) eat'	
16a Annagu waanuu	tag- naa .	16b Annagu waanuu	cun- naa
We(excl) DM-1 st per excl pl	Go-1 st p pl pres	We(exc) DM 1 st per excl pl	Eat-1 st p pl pres
'We (excl) go'		'We (excl) eat'	
17a Idinku waydin	tag- taan .	17b Idinku waydin	cun- taan .
You-all DM-2 nd per pl	Go-2 nd per pl pres	You-all DM-2 nd per pl	Eat-2 nd per sg pres
'You (all) go'		'You (all) eat'	
18a Iyagu way	tag- aan .	18b Iyagu way	cun- aan .
They DM-3 rd per pl	Go-3 rd per pl pres	They DM-3 rd per pl	Eat-3 rd per pl pres
'They go.'		'They eat.'	

Saeed (1986) analyzes -t- as the marker of second person singular, third person singular feminine, and second person plural, and -n- as the first person plural marker, while first person singular, third person singular masculine, and the third person plural have null markers. However, he shortly moves to present the paradigm as a single morpheme inclusive of person, number, gender, and tense, as the sub-division of the morpheme has no apparent consequence except in the irregular 'be' verb. Either way, composed of a null/t/n person-gender-number marker and a tense marker, or treated as a single morpheme, the forms satisfy Moravcsik's (1978) definition of agreement (15), which I will adopt.

(15) *Definition of an agreement marker* (Moravcsik 1978)

“a grammatical constituent A will be said to agree with a grammatical constituent B in properties C in language L if C is a set of meaning-related properties of A and there is a covariance relationship between C and some phonological properties of a constituent B1 across some subset of the sentences of language L, where constituent B1 is adjacent to constituent B and the only meaning-related non-categorical properties of constituent B1 are the properties C.”

Applying this definition to Somali on the data exemplified above, the verb endings always occur with the same phonological properties (following the phonological (Sandhi) rules of the language, not illustrated here) in the present tense in the same location, which is to the right of the verb, with the same “meaning-related properties” (semantic properties) as the subject pronoun. The only form which can intercede between the verb and the present tense agreement morpheme is the progressive suffix marker which is also an agreement morpheme in the tense-aspect marking (TAM) system which must occur in the same rigid location. (This is not illustrated here.) Thus, this system seems to fit the definition of agreement.

Example 19 contrasts the present tense form for singular first person with the past tense form for singular first person found in 11a. The full past tense paradigm can be found in Appendix A.

19	Aniga waan	tagay	11a	Aniga waan	tagaa
I	DM-1 st per sg	Go-1 st per sg past	I	DM-1 st per sg	Go-1 st per sg pres
	‘I went’			‘I go’	

In this example, the past tense suffix, which shows agreement for first person singular, like the present tense suffix, which shows agreement for first person singular, satisfies the definition of an agreement morpheme in precisely the same way as the present tense

suffix does. Additionally, as with the present tense suffix, the only item that can intervene between the verb and the past tense suffix is the progressive suffix, not illustrated here.

Declarative Marker

Declarative marker and Interrogative Marker.

In the sentences above, you will notice that the second word in these declarative sentences is analyzed as a declarative marker (as distinct from a topic marker, or an interrogative marker) (Saeed (1999)) An example of this is shown in 20 below. The interrogative form utilizes the particle *ma* in the second position rather than the declarative marker *waa*. The relevant constituents are highlighted in bold. The full paradigm can be found in Appendix A. (QM stands for Question Marker.)

20 Aniga **ma** tagay?
 I QM Go-past-sg-1st per
 ‘Did I go?’

18a Aniga **waan** tagay
 I DM-sg-1st per Go-past-sg-1st per
 ‘I went’

Declarative Marker and the Subject Clitic Pronoun.

The declarative marker, as seen in examples 11-18 also varies its form based on person. Saeed (1999) refers to the form after the declarative marker as a clitic pronoun which coalesces with the declarative marker *waa*. The paradigm is presented in Table 1.

Table 1: Declarative Marker Forms

Declarative marker coalesced with the ‘clitic pronoun’

waan
 DM-1sg

Declarative marker and clitic pronoun written separately

waa +aan

waad DM-2 sg	waa +aad
wuu DM-3-masc-sg	waa+uu
way DM 3-fem-sg	waa+aay
waynu DM 1-(incl)-pl	waa+aynu
waanuu DM 1-(excl) pl	waa+nuu
waydin DM 2-pl	waa+aydin
way DM 3-pl	waa+ay

Saeed does not suggest that the clitic pronouns are free morphemes; he analyses them as bound morphemes. The analysis of the coalescing particle as a clitic pronoun, however, is problematic, since the existence of two noun phrases each competing for the same theta-role would violate the first part of the Theta-Criterion repeated here:

- i) Each argument must be assigned one and only one theta-role
- ii) Each theta-role must be assigned to one and only one argument

If we theorized that one of these two pronouns existed without a theta-role, this would violate the second part of the theta-criterion which states that each argument must be assigned a theta-role. Violating the theta-criterion is extremely problematic, because this is our mechanism for ensuring that there are exactly the right number of nominal and phrases required to understand the sentence. The theta-criterion are the cornerstone of sentence construction, and for this reason are held to be essentially inviolable. Motivated

by the theta-criterion, I will attempt to reanalyze this particle as an agreement marker, rather than a clitic pronoun, later in the paper.

The critical piece of evidence for my later argument that these are agreement markers is that these forms do not in fact ever get used as independent, unbound morphemes, and they do not ever appear after or coalesced with other constituents of the sentence (for example noun, verbs, adjectives, adverbs, and so forth). It should be noted that although these forms are often written separately in formal writing, there is no evidence to suggest that they ever occur as unbound morphemes. Diriye (personal communication) says that this is a writing convention introduced in the early 70's when a written standard was being developed. Published linguistic analyses of the language, however, concur that these are bound morphemes.

Focus Markers and other Sentential Markers

In addition to the declarative marker *waa*, there are three other forms which may occur in the same position in declarative sentences. They are the focus markers *waxaa*, *baa*, and *ayaa*. The interrogative marker *ma* does not display agreement forms, but other forms which replace *waa*, *baa*, *waxaa*, and *ayaa*, such as the relative clause marker, do.

Verb Phrase

Saeed (1999) argues that in Somali “noun phrase objects do not occur in a fixed position relative to the verb,” (to rephrase with an English example, you can get word orders like “Ice cream, John likes”, “John likes ice cream” “Ice cream likes John” and so forth, however, unlike English it is problematic to say that one of these configurations

represents a base word order), but “the verbal group on the other hand may contain clitic pronouns representing any argument of the verb, including subjects, and they occur in a fixed position relative to the verb and other elements” (p. 163). (To rephrase with an English analogy, this would be equivalent to the make-believe sentences: Ice cream, John, he likes it; John, ice cream, he likes it; John, he likes it, ice cream; and Ice cream, he likes it, John; where the pronouns and verb stay in a fixed position, but the equivalent noun phrases move around.) He also argues that this is evidence that Somali does not have a verb phrase, and that full NPs in Somali appear to be adjuncts, not complements to the verb.

Saeed describes the formal structure of the Somali verb phrase as follows:

“The verbal group is a head-final phrase with dependent elements occurring as clitics preceding the head verb. Its structure may best be described by a template:

21. [vgp -S PRO O PROI-ADP-ADV1-ADVII-OPROII-V].”

Vgp refers to ‘verbal group phrase’ which is every thing to the right of the topic marker or declarative marker, including the piece I will analyze as an agreement marker-- *S Pro* which refers to the subject pronoun. *OProI* refers to the direct object pronoun. *ADP* refers to the adverbial phrase. *ADV1* refers to the first adverb in the sequence, *ADVII* refers to the second adverb in the sequence. *OproII* refers to the indirect pronoun. *V* refers to the verb. Again, to give an analogy in English, this suggests a clitic phrase such as: *-I it-to-towards-with her came*, where all of the things connected with dashes are clitics, the subject is cliticized to the declarative marker, and the *her* does not coalesce with anything, but may or may not be cliticized to the verb itself.

In the next section I will argue that in fact the fixed position subject clitics (shown coalesced with ‘waa’: *waan*, *waad*, *wuu*, *way*, *waynu*, *waanu*, *waydin*) are agreement markers and that full noun phrases, like *ice cream*, only appear to be adjuncts because they have moved out of the verb phrase into CP adjoined positions, as in *Ice cream, I like*. I will also argue that Somali is a *pro-drop* language with free inversion (see page 17: ‘has telephoned, Gianni’), and I will argue that the positioning of the noun phrases, while variable, always allows for recoverability of the noun phrase’s thematic and case relationship through the inflection on the verb. Finally, I will argue for a head-initial CP and a head-final IP.

Analysis of the data

This section will seek to analyze the subject clitic as an agreement marker, and argue for Somali as a subject *pro-drop* language. In order to avoid the violation of double filling a theta role, and thus violating the theta-criterion, pronoun-doubling clitics must be reanalyzed as agreement marking, unless they do not satisfy the definition of agreement, restated here from above.

Definition of Agreement Moravcsik (1978):

“a grammatical constituent A will be said to agree with a grammatical constituent B in properties C in language L if C is a set of meaning-related properties of A and there is a covariance relationship between C and some phonological properties of a constituent B1 across some subset of the sentences of language L, where constituent B1 is adjacent to constituent B and the only meaning-related non-categorical properties of constituent B1 are the properties C.”

Using the above definition, I will restate it, filling in A,B,C and so forth with the relevant constituents in Somali and giving an example of this constituent in parentheses. In

Somali, the full NP pronoun (John) constitutes A, the declarative marker (waa) constitutes B, the ‘clitic pronoun’ (uu) constitutes B1, and C equals person-and-number agreement. Thus, restating the definition of agreement above: The subject (John) (A) agrees with the declarative marker (waa) (B) for the phi-features person, gender and number (singular, masculine, third person) (C), because in Somali (L), person and number are meaning-related properties of the subject (John) (A) and there is a covariance^y relationship between the phi-features (singular, third person, masculine) (C) and the phonological properties of the ‘clitic pronoun’ (uu) (B1) across a subset of sentences in Somali (L), where the ‘clitic pronoun’ (uu) (B1) is adjacent to the declarative marker (waa) (B) and the only meaning-related non-categorical properties of the ‘clitic pronoun’ (uu) (B1) are the phi-features (singular, third person, masculine) (C).

Somali, like the fixed clitic languages Arabic and Hebrew, has clitics which always occur attached to the same category and their relative position is invariable. Moreover, these fixed clitics fulfill the definition of agreement. For this reason, just as it was found preferable to analyze the fixed clitics of Arabic and Hebrew as agreement markers, so too the fixed clitics of Somali are better analyzed as agreement markers.

Evidence for Pro-drop

Pro-drop is possible both with sentences using declarative markers and other sentential markers, such as focus markers and relative clause markers which have agreement on them, and with the interrogative marker which doesn’t display agreement marking. This is possible, because *pro-drop* is being licensed by the agreement marking on the verb which is always present. Examples 22 and 23 illustrate *pro-drop* with *waa*

and with *ma*. (A complete paradigm for *waa* and *ma* can be found in Appendix B.)

Column A shows a sentence with an overt NP phrase, while column B shows a *pro-drop* sentence without an overt NP phrase. The meaning of the two sentences is the same, although those in the (a) column may get a more emphatic reading. The existence of *pro* must, according to the theory described in Section 2, be assumed; otherwise, the sentences in the (b) column would be ungrammatical since the verb still has an agentive theta-role to assign, and they would thus violate the Theta-criterion.

Column A
Overt Pronoun

Declarative Marker

22a Aniga **waan** **tag-ay**
I **DM-sg-1st per** **Go-past-sg-1st per**
'I went'

Question Marker

23a Aniga **ma** **tag-ay?**
I **QM** **Go-past-sg-1st per**
'Did I go?'

Column B
Dropped Pronoun

22b **Waan** **tag-ay**
DM-sg-1st per **Go-past-sg-1st per**
'(I) went'

23b **Ma** **tag-ay?**
QM **go-past-sg-1st per**
'Did (I) go?'

Analyzed as *pro* sentences, the underlying structure of sentences 22b and 23b would be 24 and 25 below, where *pro* is the non-overt subject.

24 *pro* **waan** **tag-ay**
(I) **DM-sg-1st per** **Go-past-sg-1st per**
'(I) went'

25 *pro* **ma** **tag-ay?**
(I) **QM** **go-past-sg-1st per**
'Did (I) go?'

If the researcher is unaware of the existence of *pro*, then in the sentences above, the most natural overt element to fill the position of subject is the subject agreement marker on the declarative marker *waa* and focus markers *waxaa*, *baa* and *ayaa*; however,

sentence 23 occurs without the subject clitic, which leaves the alternative analysis of having (1) a sentence without a subject, or, alternatively, (2) having a non-overt subject. Since not having a subject in sentence 23 would violate the theta-criterion which require that all theta-roles be assigned, this is unsatisfactory, and thus the option of positing a non-overt subject which receives the subject theta-role is preferable, because violating the Theta Criterion leaves us able to prevent the generation of too few or too many noun phrases in a sentence.

Let us now turn to examining the requirements to license a non-overt pronoun. One of the requirements for a language to license *pro* is that pronouns must be recoverable from the context. Specifically, *pro* is licensed by an overt (rich) Agr category which is co-indexed with it. This means that in agreement-rich languages like Spanish and Italian which inflect their verbs for person, number, and gender *pro*-drop is possible, but in languages like English which are agreement poor it is not. Somali is clearly an agreement-rich language with similar verbal inflection. There is subject, number, and person agreement on both the verb and declarative marker. Thus, the phi-characteristics of the pronoun are recoverable via the agreement features on both the declarative marker and the verb which is co-indexed to it. Therefore, *pro* can be licensed in Somali, just as it is in Spanish and Italian.

Somali exhibits another characteristic of *pro-drop* languages observed by Taraldsen (1981) and Rizzi (1982). Among these characteristics is a property called *Free Inversion*. Free inversion is a process in which the overt subject of a finite clause can appear freely in the clause-final position following the verb phrase and other VP-material. An example from Italian is shown in 10, repeated below:

26. Ha telefonato Gianni
 Has telephoned Gianni
 ‘Gianni has telephoned’ (Ouhalla, 1999, p. 314)

Such a noun phrase is clearly an adjunct to the sentence. The free inversion feature observed in Italian can potentially be seen in the following Somali sentence in (27).

27. Moos waa cunay anigu.
 Banana DM eat-past-1st per sing I
 ‘I ate a banana.’

(27) could be analyzed, as Saeed (1999) analyses it, as a non-fixed, adjunctive position for full-NPs in a system without rigid word order for full-NPs. However, it could also very well be an example of *free inversion*, an analysis which explains some of the apparent examples of flexible word order in Somali, without suggesting that Somali does not have a verb phrase. Possible discourse properties to be investigated which might be associated with this position are suggested in the discussion at the end of the paper.

Constituent Locations: Head-final Head Parameter of IP

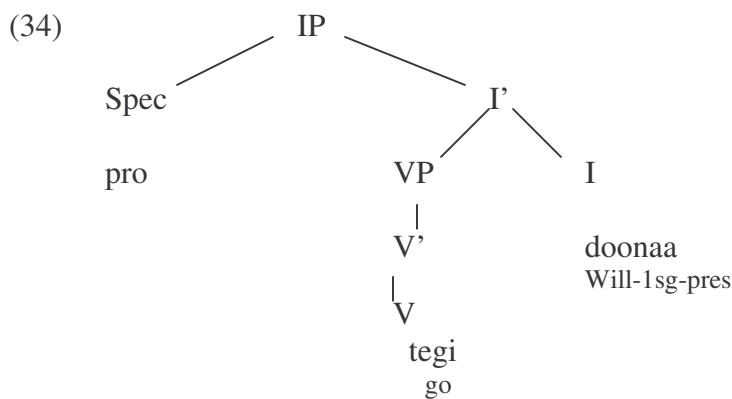
Somali appears to have a head-final head parameter setting for IP since the auxiliary verb, which is the head of IP, appears to the right of the verb as shown in the sentences below (28-33), as do the tense and aspect markers in the sentence previously analyzed.

28. Way tegi **lahaayeen**
 DM-3-plgo+infinitive have-would
 ‘They would go/they would have gone’

29. Way tegi **doonaan**
 DM-3-plgo-infinitive will-3-pl
 ‘They will go.’

30. Way tegi **jireen**
 DM-3-pl go+infinitive used to-3-pl
 'They used to go'
31. Waan tegi **karaa**
 DM-1-sg go+infinitive can-pres-1-sg
 'I can go'
32. Waan tegi **waayey**
 DM-1-sg go-infinitive fail-past
 'I failed to go'
33. Waan tegi **gaadhay**
 DM-1-sg go-infinitive just-fail
 'I almost went'

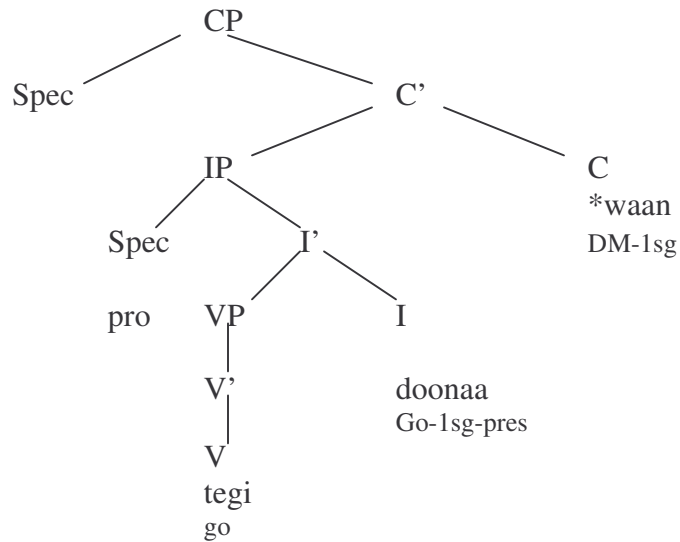
These examples suggest a head-final clause structure for I, as seen in (34):



'(I) will go.'

Theorists would like all phrases to have the same parameter setting since it permits a broader generalization. If this were the case, then CP would also need to be head final, as illustrated in (35). This structure, however, is problematic, since the head of C can be seen to the right of IP and not to the left of IP, putting the interrogative and declarative markers on the wrong side of the verbal construction.

(35)



‘(I) will go.’

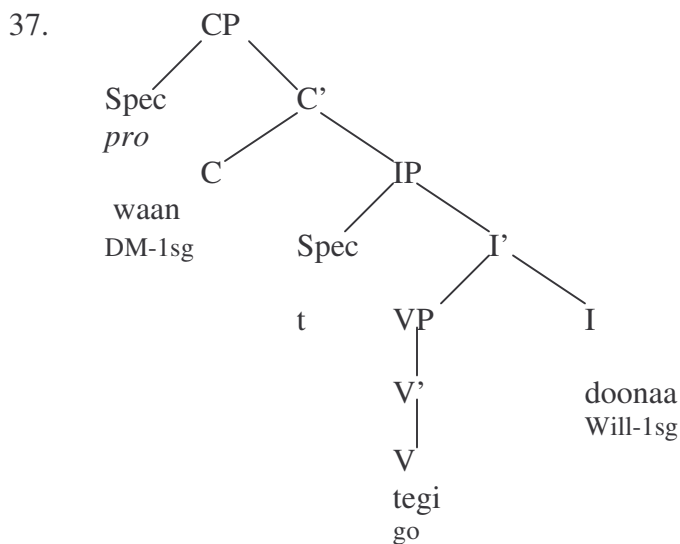
One way to save this analysis might be to place the declarative and interrogative markers in the specifier position of C. This is problematic for several reasons, however. First, declarative and interrogative markers are normally considered head categories, and this would be inconsistent with the theory and cross-linguistic observations. Second, relative clauses utilizing *in* ‘that’, have their complementizer head initially, as seen in example 36, showing that the head of CP is initial, despite IP being head final.

36. Waxaan doonayaa inaan tago
DM+1sg want+prog+pres that+1sg go+pres-gen
I want to go (lit: I want that I go).

Third, if the declarative and interrogative markers were in the specifier position, there would be no landing site for *wh*-phrases short of projecting another CP above the original CP, yielding 2 empty C heads after the verb. Fourth, while NPs occasionally show up sentence finally, the *free inversion* concept gives us a potentially better explanation for this as an adjunction structure than supposing that somehow they are able to move right

into a head position from a specifier (phrasal) position, which would also be inconsistent with cross-linguistic observations and the theory.

The better structure which would conform to the evidence of the language would make CP head-initial, while IP is head-final. A split head parameter (meaning CP is head-initial, while IP is head-final) is attested in several languages, such as German. Using a split head parameter analysis allows us to correctly locate the declarative marker and interrogative markers to the left of the verb. As you can see in the structure in (37), *waan* now ends up on the correct side of the VP phrase.



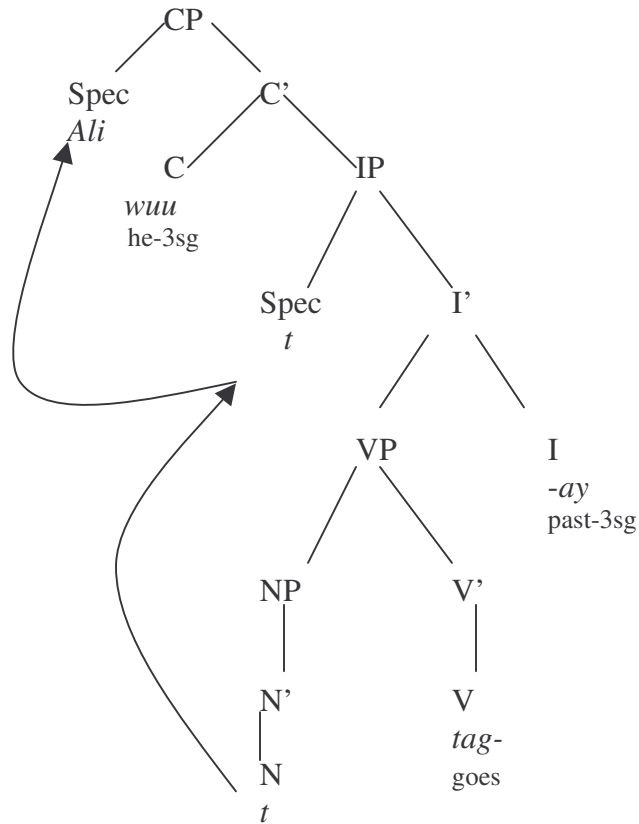
‘(I) will go.’

Subject movement out of the verb phrase

Many theories of generative grammar now analyze now analyze the subject position as DP/NP movement out of the verb phrase, where it receives a thematic role, to the specifier of IP where case is assigned. Thus, as in English, the subject in Somali is

analyzed as moving from Spec-VP to Spec-IP to get case. In contrast with English, however, the Somali subject continues moving up to Spec-CP. 38 illustrates this movement pattern.

38.



'Ali wuu tagay.'
Ali went.

On the surface, this appears similar to German, however, one possible explanation for this movement is that the Somali subject moves to Spec-CP, because it must match its phi-features with the head of CP (something which does not seem to pertain to German). This explanation offers a lot of possibilities, however, one potential problem for this

analysis comes from focused subjects. In focused subjects, such as 39 below, no agreement takes place between the subject and the sentential marker.

39. Aniga waa cunay.
I DM eat-1st person singular past tense
I ate (it).

This is unexpected, since the subject is in the specifier of CP. One might also expect to get instead the sentence in 40 in instances where there is no agreement with the sentential marker, but 40 is ungrammatical.

40. *waa aniga cunay.
DM I eat-1st person singular past tense
'I ate (it).'

These facts suggest that the agreement on the sentential markers may be more complicated than verbal agreement, requiring certain discourse facts to pertain. Since this is a clearly definable subset of sentences, and our argument above generally pertains, we might be allowed to temporarily set this issue aside to be taken up in future work that attends to the information structure of Somali sentences, and to the newly evolving field of the pragmatic-syntactic interface (Ward 1999).

Conclusion

This analysis provides evidence that Somali is a *pro-drop* language, and that the bound morpheme on the sentential markers is an agreement morpheme, not a clitic pronoun. There is strong evidence that Somali is a *pro-drop* language due to the

alternation of forms with and without an overt pronoun which it demonstrates, and due to the fact that it conforms to the licensing requirements of the *pro-drop* parameter. The bound morpheme on the various sentential markers (*waa, waxaa, baa, ayaa, inaa* etc) conforms to the definition of agreement, and would violate the Theta-criterion if it were analyzed as a reduplicative pronoun. Additionally, word order evidence suggests that IP is head-final, while CP is head-initial.

Pedagogical Implications

This analysis suggests that the reason Somali speakers learning English systematically omit subjects in their writing is for the same reason Spanish speakers do—their L1 is a *pro-drop* language and this is a transfer error. Thus, “walked to school yesterday” with the meaning “I walked to school yesterday” is perfectly grammatical in Somali, where agreement on the verb makes recovery of *pro*’s identity as ‘I’ completely possible. While most L2 learners learn the SVO word order of English quite quickly, Somali and Spanish speakers generally need additional instructional intervention to actually fill-in the subject position with an overt pronoun late into their L2 development, and this difference between English and Somali often needs to be pointed out explicitly for several years before the L1 pattern is subordinated to the English rule requiring a subject when writing in English.

Additional evidence for this analysis could be gathered by collecting Spanish L1 and Somali L1 writing samples and comparing them to writing samples of students from non-*pro-drop* languages, such as Russian. It would also be useful to look at what types of instruction help students internalize the obligatory subject rule in English. Does

specific correction of this error reduce its frequency? It would also be useful to look at the effectiveness of correction on transfer errors, as compared to the effectiveness of correction on errors related to various stages of acquisition. Does correction really facilitate learning when the errors are due to transfer, but not facilitate learning when it is due to being a stage of universal language acquisition?

Theoretical Implications of Pedagogical Observations

This research also raises an interesting theoretical question. *Pro-drop* is thought to be licensed by the rich agreement in *pro-drop* languages, so one would think that the surface pattern would not simply transfer to English which is agreement poor and can't license *pro-drop* generally. One must wonder if the licensing requirements have anything to do with the production of the structure at all, or if they are really receptive constraints that play out in the interaction of the principle of least effort on the part of the speakers and the hearers. More specifically, from the perspective of the speaker, the principle of least effort would suggest that all words should be of the form "aaaaa". Of course, from the perspective of the hearer, the principle of least effort suggests that the phonology and syntax should be maximally differentiated in order to easily understand what is being said. The interaction of these two principles causes languages to develop (and change) in particular ways. Thus, the principle of least effort on the part of the speaker would encourage *pro-drop*, but the principle of least effort on the part of the hearer prevents *pro-drop* from being acceptable unless the information is otherwise recoverable from the context. That L2 speakers do not make this transition is quite fascinating. It might suggest that prior to some threshold the cognitive load is simply too

high to attend to the needs of the hearer. Alternatively, it might suggest that although there is an identifiable principle that is created by the constraints of least effort that holds across languages for practical reasons, in fact speakers are simply learning the patterns and facts of their language without attention to the licensing constraints. Restated, it could be that the licensing constraints have no psychological reality in deciding what is a permissible structure or not, and that they are only the residual effect of processing constraints that play out in the give-and-take of both the speaker and hearer seeking to communicate in speech with communally defined “least effort”. The actual meaning of “licensing” clearly remains to be investigated, and looking at the violation of “licensing” requirements by L2 speakers would be an interesting endeavor.

Implications of Theory on Practice

The above observations on licensing also suggest that in addition to looking at error rates between students with *pro-drop* L1s and students with non-*pro-drop* L1s, we might also want to look at differences between students who license *pro* differently in their L1s. Are there differences in the types of errors produced, the effectiveness of correction, and the rate of acquisition of the overt subject when students come from languages like Spanish and Somali which license *pro* via agreement marking on the verb, and students who come from language backgrounds like Japanese where *pro* is licensed by level of activation in the discourse. Finally, what do students from languages like German and English do, who have *pro* licensed by illocutionary force in imperatives when forming imperatives in languages which do not license *pro* in this environment.

Theoretical Issues for Further Investigation

In his 1999 paper on the properties of free inversion structures in Italian, Ward discovers that Italian, which lacks an existential there construction, but has a presentational there construction *ci*, uses the post-positioning of subjects to place discourse new information later in the sentence, and all of these examples of subject postposing seem to be generated in order to make the sentences more harmonic with regard to preferred (unmarked) discourse structure. Future research on Somali needs to investigate the information structure role on word order variation.

Related to the need to establish a clearer understanding of the syntactic-pragmatic interface, is a need to clean-up CP, which has been a dumping ground for context-dependent grammatical structures. Among the structures dumped here are adjunction of thematically governed structures which have moved due to requirements of the pragmatic system, items denoting illocutionary force [Q] being a prime example, relative pronouns, complementizers, and so forth. If we wish to improve the accuracy of our grammatical models, we are going to need to begin looking at the structure of the context-dependent grammatical items in CP, evolve this, and look at its relationship to the context-free components below CP.

References

Andrzejewski, B.W. 1961. "Notes on the substantive pronouns in Somali." Reprinted from *African Language Studies II*, 1961. London: School of Oriental and African Studies, University of London.

- Andrzejewski, B.W. 1964. *The declensions of Somali nouns*. London: School of Oriental and African Studies, University of London.
- Andrzejewski, B.W. 1975a. "The role of indicator particles in Somali." Monographic *Journals of the Near East, Afroasiatic Linguistics*. 1,6.
- Andrzejewski, B. W. 1975b. *Indicator particles in Somali*. Malibu: Undena Publications.
- Andrzejewski, B.W. 1979. *The case system in Somali*. London: School of Oriental and African Studies, University of London.
- Andrzejewski, B.W. 1974. "The dichotomy between extensive and restrictive verbal paradigms in Somali and its parallels in Oromo." Reprinted from *Atti del Secondo Congresso Internazionale di Linguistica Camito-Semitica Firenze*. 16-19 April.
- Andrzejewski, B.W. 1984. "The role of accentual patterns in subject/object differentiation in Somali and its parallels in Paraniotic languages." Reprinted from *Current Progress in Afro-Asiatic Linguistics, Papers of the 3rd International Hamito-Semitic Congress*. James Bynon, ed. Amsterdam; Philadelphia: John Benjamins.
- Dik, S. C. 1989. *The theory of functional grammar*. Providence, RI: Foris Publications
- Diriye, M. 2001. *Le Somali, dialects, et histoire*. Department of Linguistics, University of Montreal.
- Lamberti, M. 1988. *Die Nordsomali-Dialekte: Eine synchronische Beschreibung*. Heidelberg: Carl Winter Universitaesverlag.
- Moravcsik, E. 1978. Agreement. In J. Greenberg Ed. *Universals of human language: Volume 4: Syntax*. 331-374. Stanford, CA: Stanford University Press.
- Orwin, M. 1995. *Colloquial Somali: A complete language course*. London: Routledge.

- Ouhalla, J. 1999. *Transformational grammar*. London: Arnold.
- Mansur, A. O. & Puglielli, A. 1999. *Barashada naxwaha af Soomaaliga: A Somali school grammar*. London: Haan Publishing.
- Rizzi, L. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, L. 1986a. Null objects in Italian and theory of *Pro*. 501-558. *Linguistic inquiry*. 17.
- Rizzi, L. 1986b. On the status of subject clitics in Romance. In O. Jaeggli & Silv-Corvalan C. (Eds). *Studies in Romance linguistics*. 391-419. Dordrecht: Foris.
- Roberts, I. & Shonsky, U. 1995. Pronominal enclisis in VSO languages. In R. Borseley & I. Roberts (Eds), *The syntax of the Celtic languages*. 171-199. Cambridge: Cambridge University Press.
- Ruhlen, M. 1991. *A guide to the world's languages: Volume 1: Classification*. Stanford, CA: Stanford University Press.
- Saeed, J. 1999. *Somali*. Amsterdam/Philadelphia: John Benjamins.
- Taraldsen, K. (1981) The theoretical interpretation of a class of marked extractions. In A. Belletti, L. Brandi & Rizzi, L. (Ed). *Theory of markedness in generative grammar*. 475-516. Pisa: Scuola Normale Superiore.
- Ward, Gregory. 1999. "A comparison of postposed subjects in English and Italian," in *Functions and structure*, edited by Akio Kamio and Ken-ichi Takami. 3-21. Amsterdam/Philadelphia: John Benjamins.
- Van Vallin, Robert D., Jr. 2001. *An introduction to syntax*. Cambridge, UK: Cambridge University Press.
- Zorc, R. D. & Issa, A. A. 1990. *Somali textbook*. Wheaton, MD: Dunwoody Press.

ⁱ LFG, for example, does not instantiate a concept of *pro*, but rather instantiates the phenomena as a lack of extension of a specifier for IP in its c-structure, and a contribution of thematic material from agreement

markers in its f-structure. That said, it is not actually clear to me how LFG deals with *pro-drop* in English commands since English does not have agreement marking available to contribute thematic material in f-structure, whereas Government and Binding Theory has only to toss the issue into pragmatic licensing, and HPSG can deal with the issue in a number of ways. Likewise, *pro-drop* in Japanese seems to have its thematic material supplied by discourse structure, not sentential structure, and this suggests that, as VanVallin and LaPolla (1997) point out, we are working out our syntactic theories on too narrow a field of operation. To provide a uniform account of the licensing of *pro-drop*, it seems that we are going to have to expand the scope of our syntactic analyses to the discourse level. These issues seem to be over-looked, however, as a general rule by the field of syntax, as we do not yet seem ready as a field to take the next step into understanding the syntactic levels of discourse, whereby VP represents syntactic levels of semantics related to thematic structure of verbs, IP represents syntactic levels of semantics related to time and telicity, and CP currently lumps everything else together, representing markers of mood with open propositions, such as WH-questions, other discourse marked structures, and complementizers which exert control over the entire sentence.

ⁱⁱ Subsequent work has argued that *pro* is licensed in command forms in many languages without agreement morphemes, such as English, and that Japanese utilizes and licenses *pro* through discourse insertion. The evaluation of other means of licensing *pro* is not undertaken in this paper since the original theory is satisfactory without modification, for the analysis presented here, and since it does not affect the analysis in this paper in any way, it is not evaluated here, however there is likely more breadth to what may license *pro* than what is presented here.

ⁱⁱⁱ This principle gets us *pro* in specific contexts, such as English Commands, and in discourse oriented languages like Japanese.

^{iv} I.e. not English which can only license *pro-drop* in commands

^v This term may have various meanings, but is used since it is in the definition above.