

Formal Analysis of Some Aspects of Amharic Noun Phrases

Sisay Fissaha Adafre
Institute for Logic, Language and Computation
University of Amsterdam
Nieuwe Achtergracht 166
1018 WV Amsterdam
The Netherlands
[sfissaha@science.uva.nl]
<http://www.illc.uva.nl>

Abstract. This paper presents a syntactic analysis of a fragment of Amharic noun phrases. Amharic noun phrases possess some peculiar characteristics especially relating to the syntax of the definite article. The definite article shows some morphological and distributional characteristics distinct from languages such as English and German, which makes it difficult to directly apply the methods proposed for these languages to the analysis of Amharic noun phrases. In this paper, we provide a formal analysis of Amharic noun phrases by combining the methods of the different approaches suggested for analyzing noun phrases. The result shows that the affixial treatment of the Amharic definite article better explains the facts regarding the syntax of Amharic noun phrases.

1. Introduction

Amharic belongs to the Semitic language family and is one of the most widely spoken languages in Ethiopia. Amharic has its own script that is borrowed from Ge'ez, another Ethiopian Semitic language (Leslau, 1995:1). Amharic has a complex morphology. Thus the need for automatic processing of Amharic words is getting increasing attention, which is reflected in the recent rise in the number of Amharic language processing applications (Nega, 1999; Abiyot, 2000; Mesfin, 2001; Worku, 1997). Regarding the syntax of Amharic, except for a prototype experiment done on a Probabilistic Context Free Grammar for Amharic by Atelach (2003), very little work has been done to formalize the syntax of Amharic especially using unification based grammar formalisms. On the other hand, syntactic analysis forms the core of most Natural Language Processing applications such as Machine Translation.

The current work is part of a larger project that aims at integrating Amharic into the CAT2 machine translation system (Haller, 1993). Previous works mainly focus on the lexical and the morphological aspect of the language (Sisay, 2003a; Sisay, 2003b). In this paper, we provide a syntactic analysis of a fragment of Amharic noun phrases. In the next section, we review some of the approaches adopted

in the analysis of noun phrases. This is followed by a descriptive analysis of Amharic noun phrases, which we present in Section 3. In Section 4, we describe the main features of the two approaches, i.e. *Determiner Phrase Analysis* and *Noun Phrase Analysis*, by taking Amharic noun phrases as a case. This provides the necessary methodological basis for Section 5 where we present an analysis of Amharic noun phrases using the CAT2 formalism. In Section 6, we mention the relevance of our analysis to Machine Translation. Finally, Section 7 presents a summary of the paper.

2. Background

Several approaches have been adopted in describing the noun phrases of a language. These approaches differ on a number of aspects. The most significant area of controversy, which resulted in two major streams, is the one concerning the head of a nominal phrase. Approaches assuming a determiner as head of a nominal phrase are called *Determiner Phrase Analysis* (DP Analysis) (Abney, 1987) whereas those adopting a noun as head of a nominal phrase are called *Noun Phrase Analysis* (NP Analysis) (Pollard and Sag, 1994).

The DP analysis usually stems from the need to provide an analysis of nominal phrases that

parallels clause structures. A number of morphological, syntactic and semantic arguments have been provided in support of DP analysis. Some of the supporting evidences include; the parallelism between agreement features of noun phrases and clauses; the similarity between the argument structure of verbs and the corresponding nominalizations (e.g. destroy vs. destruction); the distribution of adjectives in noun phrases and adverbs in clauses; and the fact that noun phrases need to be functionally complete when used as an argument in a way similar to the matrix clause (Bernstein, 2000:538-544).

On the other hand, the NP analysis remains the main theoretical claim of grammatical theories such as HPSG (Pollard and Sag, 1994). Here again several supporting evidences have been provided. Other approaches try to show that pure DP or NP analyses do not provide an adequate explanation for the phenomena found in a number of languages such as German (Netter, 1996:111).

Different proposals have been made in the description of Amharic noun phrases. Girma (1994:93) provides a KP (Case Phrase) analysis of Amharic noun phrases. He classifies functional categories in Amharic into strong and weak based on their surface realization. According to Girma (1994:94), weak functional heads are realized as bound morphemes. On the other hand, strong functional heads are unbound morphemes. For example, the Amharic definite article *-u* is a suffix morpheme and hence is considered a weak functional head (1). On the other hand, demonstratives such as *yh* ('this') are independent morphemes that appear in prenominal positions and, therefore, are strong functional heads (2) (Girma, 1994:81-88).

1. bet-u-n
house-DEF-ACC
'the house'
2. yh-n bet
this-ACC house
'this house'
3. wädä bet-u
PREP house-DEF (HIS)
'to DEF (HIS) house'

Girma (1994:83) also treats case as a functional head. The Amharic accusative case marker *-n* is a suffix morpheme and a weak functional head. Other case markers, especially those, which are treated as prepositional morphemes in traditional

grammar, e.g. *wädä* ('to') (3), are either independent morphemes appearing prenominally or bound prefix morphemes, e.g. *kä-*. In order to explain the apparent variation in word order of elements belonging to the same category, i.e. definite article (postnominal) vs. demonstratives (prenominal) or accusative case markers (postnominal) vs. other case markers (prenominal), Girma (1994) proposed the following structural configuration and the Head-to-Head movement operation.

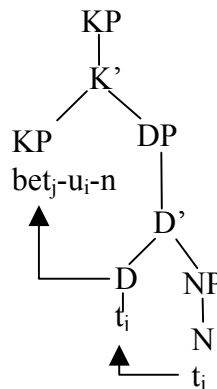


Figure 1 Head-to-Head Movement Analysis of Amharic NPs (Girma, 1994:95).

The Head-to-Head movement is applied to noun phrases that do not have modifiers. A weak head attracts a strong head as it does not have an independent surface realization, which results in the iterative upward movement of a strong head as shown in Figure 1. In case the noun phrase contains a modifier, Spec-to-Spec movement strategy is applied. In the Spec-to-Spec movement, the modifiers are moved instead of the head noun since it is the modifiers that host the weak functional heads.

Some comments are in order. The assumption that both prepositional morphemes and the accusative case marker *-n* belong to the same class is not well established. In most Amharic literature, prepositional morphemes are treated as belonging to the separate class of prepositions different from the direct object marker *-n*. Furthermore, the treatment of the Amharic definite article as an independent syntactic unit is not well motivated either, since it shows affixal properties that support its treatment as a feature of nouns, along with categories like number or gender rather than as an independent syntactic unit. On the other hand, in the analysis provided by Baye (1986), the noun is the head of the noun phrase. According to his analysis, the different constituents of a noun phrase are specifiers, i.e. determiners, possessive

adjectives and quantifiers, modifiers, complements, and the head noun.

As shown in the above discussion, there are different views concerning the syntax of Amharic noun phrases. In this paper, we provide a reanalysis of Amharic noun phrases based on the ideas developed within the framework of contemporary syntactic theories.

3. Descriptive Analysis of Amharic Noun Phrases

An Amharic noun phrase typically consists of determiners, quantifiers, adjectives, relative clauses, prepositional phrases, and a noun. These constituents appear in prenominal positions since Amharic is primarily a head final language. Determination in Amharic may be expressed using different lexical elements, such as articles or demonstrative adjectives. In addition to definiteness, the definite article also encodes gender. Regardless of gender, the form of the definite article for plural nouns is *-u*, and is identical to the form of the definite article for singular masculine nouns ending in a consonant.

Indefinite nouns are normally not marked (4). The cardinal *and* ('one') is occasionally used to indicate the meaning of indefiniteness, especially in introducing a human discourse entity, and appears preminally as shown in (5). It is used also to individuate the nominal blocking thus generic interpretation. Nouns without an associated article are sometimes ambiguous between an indefinite (4) and a generic interpretation (6).

4. wānbār amāṭṭa
chair bring-perf-HE
'He brought a chair'
5. and-it lg
'a (one)-FEM girl'
6. wānbār y-sār-all
a chair HE-made-AUX
'He makes chairs.'

Demonstratives may either head a noun phrase (as demonstrative pronoun) or may serve as determiners (as demonstrative adjectives). The demonstrative adjectives are inflected for number and gender. They also take prepositional prefixes and case markers. Unlike the definite article, demonstratives are unbound morphemes and appear preminally (7).

7. yh tqur wānbār
this-MASC-SING black chair
'this black chair'

A demonstrative adjective can not be used with a noun having the *-u* suffix. In such cases, the definite article and the demonstrative are mutually exclusive (8).

8. *yh wānbār-u
this-MASC-SING-DEF chair-DEF
'this chair'

However, there are cases in which the definite article and the demonstrative adjective co-occur in a noun phrase resulting in over-determination. For example, if an attributive adjective intervenes between the demonstrative adjective and the noun, then the attributive adjective optionally takes a definite article (9).

9. ya tlq-u wānbār
that-MASC-SING big-MASC-SING-DEF
chair
'that big chair'

Possessive expressions in Amharic are rendered by using possessive suffix pronouns attached to a head noun as in (10), or prefixing *yä* to the personal pronouns or nouns indicating the possessor (11) and (12).

10. bet-e
house-MY
'my house'
11. yä-ne bet
POSS-I house
'my house'
12. yä-šufer-u bet
POSS-driver-DEF house
'the driver's house'

The definite article suffix and the possessive suffixes are mutually exclusive and cannot appear on the same noun. However, if a noun with a possessive suffix pronoun is modified by an attributive adjective, the attributive adjective takes a definite article (13).

13. tlq-u bet-e
house-DEF house-MY
'my big house'

The direct object marker *-n* has a similar distribution to that of the definite article. If prenominal modifiers like attributive adjectives, relative clauses, or possessives precede a noun, the marker *-n* is attached to the prenominal modifiers. The direct object marker is suffixed after the definite article suffix morpheme (cf. 1). If all prenominal modifiers of a direct object noun phrase are definite then they should all carry the direct object marker as shown in (14). A direct object marker may be applied to an indefinite noun phrase. It renders a generic meaning (15).

14. ya-n tlq-u-n dmät
that-ACC big-DEF-ACC cat
'that big cat'
15. ktbat bāšta-n lä-mākälakäl y-rädal
vaccination disease-ACC PREP-protect -
HE-helps
'vaccination helps to protect from disease.'

In general, among the syntactic features that are relevant for internal and external agreement relation of noun phrases are gender, number, case and definiteness. However, the definiteness agreement relation is different from number and gender agreement. In the former case, the prenominal elements tend to specify the form of the noun in terms of definiteness. Furthermore, these agreement features are distributed among the prenominal constituents and the noun. As a result, even though the noun is the semantic nucleus of the noun phrase, syntactically, however, other constituents are equally important in determining the overall syntactic behavior of the noun phrase.

4. NP or DP Analysis?

In this section, we will have a closer look at the two views concerning the syntax of noun phrases: NP analysis and DP analysis. In doing so, we provide a general description of the approaches and try to relate them to the characteristics of Amharic noun phrases.

4.1 NP Analysis

In the NP analysis paradigm, the noun functions as the head of the noun phrase whereas determiners are usually treated as specifiers of the noun. Adjectives are treated as modifiers that select for the type of constituent they modify. The noun should therefore provide the morphosyntactic features that are involved in the agreement relation. One major argument in favor of this

analysis is that the definite article, accusative case marker, and plural marker are suffix morphemes that attach to the noun. Since definite article also carries the gender feature, NP analysis looks more attractive especially for those noun phrases consisting only of a single noun constituent (cf. 1).

However, the discussion in the previous section suggests that prenominal constituents encode much of the morphosyntactic information relevant for the external agreement of the noun phrase. This means that morphosyntactic features should percolate up the projection not only from the head noun but also from other constituents of the noun phrase. In NP analysis, particularly in HPSG Grammar formalism, the relation between the head and the different constituents within the noun phrase has been formalized into a set of general principles such as head-specifier, head-modifier, and head-complement rule (Pollard and Sag, 1994; Sag and Wasow 1999). The task of making the agreement features available in the head noun is achieved in Sag and Wasow (1999:92) through a Nominal SPR (specifier) Agreement constraint that unifies the agreement feature of the head noun with that of the specifier. The head feature principle, in turn, percolates the head features to the higher-level projections.

The relation between attributive adjectives and the head noun is treated as a head-modifier relation in which the adjective selects the noun through the head feature *mod*. The noun functions as syntactic head. In the case of indefinite noun phrases containing an attributive adjective and a noun, the head-modifier rule licenses the NPs in (16) and (17):

16. tlq dmät
big-SING cat-SING
'big cat' ('a big cat')
17. tlalq dmät-oc
big-PLU cat-PLU
'big cats'

The relevant morphosyntactic information in these constructions is number. Both adjectives and nouns agree in number. This can be handled by including the relevant agreement feature in the *mod* feature. As regards definite noun phrases, the attributive adjective also carries the definite article, number, and gender features as well as the direct object marker (accusative case). The fact that the Amharic definite article is a bound morpheme raises some problems in applying the above

approach. The problem specifically relates to the status of the definite article, i.e. whether it is an affixal morpheme, a clitic, or a word. Bulgarian (Popova, 2000) and Hebrew (Wintner, 2000) are other languages that exhibit similar morphosyntactic phenomena with respect to the definite article. Both Popova and Wintner approach the problem by first establishing the fact that the definite article in the respective languages is an affix morpheme according to criteria outlined in Zwicky (1985). The most convincing argument is the morphophonological criterion, which states that affixal elements exhibit idiosyncratic morphophonological phenomena, i.e. they are affected by the phonological properties of the host. This is also true for the Amharic definite article since the form of the definite article depends on the endings of the host.

After establishing the affixal nature of the definite article, Wintner (2000) provided his analysis of the definite article in the HPSG framework. Wintner (2000) treated the definite article as a definite marker and considered it as a head feature different from agreement features such as number and gender. In Hebrew, an adjective and a noun agree in the definiteness feature hence the *def* feature of the adjective is co-referenced with the *def* feature of the noun. Unlike in Hebrew, adjectives and nouns in Amharic show a different agreement relation. In Amharic the definite adjective selects a noun which is negatively marked in terms of definiteness *def=no*.

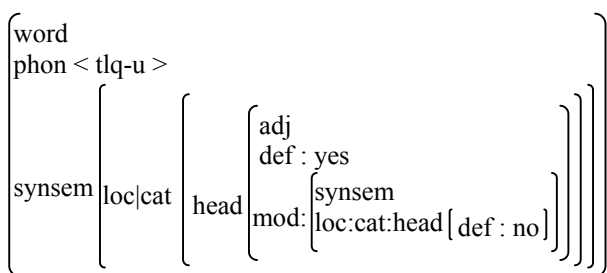


Figure 2 Lexical Entry for Definite Adjectives

Since in the NP analysis, the noun is the head of the phrase, the head-feature principle copies the feature of the noun to the mother node rendering the whole phrase indefinite in case the head noun is generic or in the plural form. In order to have a correct output, the definiteness feature should percolate from the adjective to the mother node. However, the head-feature principle does not percolate the head features of the modifier.

In general, the treatment of the definite article as an affix and its incorporation as a head feature of a

noun is an important aspect of the above analysis, which we have adopted in our analysis of Amharic noun phrases. However, the lack of a typical agreement relation (with respect to definiteness) similar to number agreement between the pronominal elements and the head noun redefines the relation among the pronominal elements and the head noun as selection rather than agreement. As we have seen in previous paragraphs, the pronominal elements tend to specify the properties of the noun. This points to the need to look at alternative approaches for the description of Amharic noun phrases. Therefore, in the next section, we examine in more detail the determiner-based (DP) analysis.

4.2 DP Analysis

In DP analysis, the determiner serves as the head of the noun phrase. Although nouns constitute the semantic core of noun phrases, syntactically, determiners also play a significant role. They provide important morphosyntactic features that are relevant to both the internal and the external agreement relations of noun phrases. DP analyses are the main approaches within Chomskian paradigm. As the current work focuses mainly on non-transformational and lexical based approaches, transformational accounts of the problem mentioned will not be considered.

After a detailed analysis of the pros and the cons of pure NP- and DP-based analysis, Netter (1996) provided a DP-like analysis of German noun phrases within the HPSG formalism. Netter (1996:122) organized the head features into groups of *major* and *minor* features. The major feature includes the category and agreement features while the minor feature includes features that are used to control the specification of the maximal projection.

Netter (1996:121) classified the German lexicon into substantive and functional categories. The substantive category includes among others nouns, verbs, and adjectives whereas the functional category consists of functional elements such as determiners and quantifiers. He introduced the concept of functional completeness that refers to the requirement that phrasal projections need to have the necessary functional marker to form their maximal projections.

Netter (1996:128) uses the binary feature *fcompl* in order to mark functional completeness. Singular count nouns are marked in the lexicon as functionally incomplete, i.e. *fcompl=no*. Plural nouns and mass nouns are underspecified with

respect to functional completeness. Determiners like definite articles are heads of the maximal projection, and select nouns which are functionally incomplete, *fcompl=no* (cf. Figure 3). In order to control the word order of the elements in prenominal position and to specify the intermediate projection at which prenominal modifiers like adjectives may attach, he used another binary feature, *spec*. Adjectives select through the *mod* feature a noun which is marked as *spec=no*. Since the head of the intermediate projection is the noun, the mother node will have the feature *spec=no*. A determiner is marked in the lexicon as *spec=yes*. A determiner selects a noun or an intermediate projection which is marked as *spec=no*. In this case, since the head is the determiner, the mother node will be specified as *spec=yes*. This prevents multiple occurrences of determiners in a noun phrase. The *fcompl* and *spec* features are minor features.

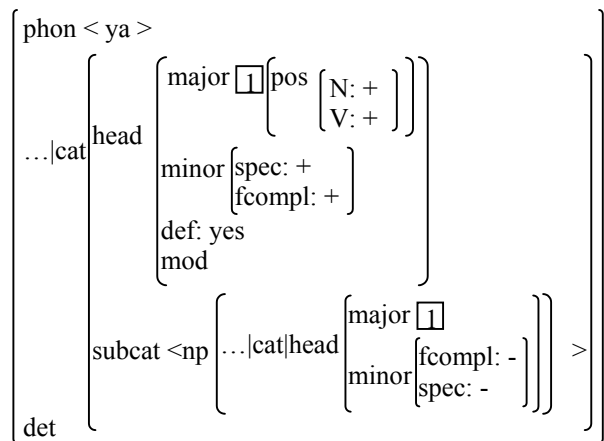


Figure 3 Lexical Entry for Demonstrative adjectives

Unlike German or English, Amharic singular count nouns do not necessarily require articles to form their maximal projection. Therefore, the requirement that singular count nouns be inherently functionally incomplete does not apply to Amharic nouns since Amharic singular count nouns, like plural or mass nouns, are underspecified with respect to functional completeness. Furthermore, Amharic has a bound definite article morpheme, the analysis of which is not obviously covered in the work of Netter since German does not exhibit such a phenomenon. Therefore, there are some constructions, which make it difficult to directly apply the proposal made by Netter. For example, nouns with definite article suffix do not allow other specifiers or modifiers, such as demonstrative adjectives or attributive adjectives carrying the definite article

suffix. They may be considered as functionally complete, i.e. *fcompl=yes*, and specified, i.e. *spec=yes*. This gives an impression that the definite article suffix has the same interpretation as the definite article of German. However, the definite article suffix does not impose the same constraint when it appears on an attributive adjective.

18. tlq-u dmät
big-DEF-MASC dmät
'the big cat'

The NP in (18) contains definite article suffix attached to the attributive adjective. This, however, does not imply that the resulting noun phrase is functionally complete and specified. Instead, the NP in (18) has properties similar to generic or plural nouns. It may form a maximal projection by itself or may further combine with a specifier, for example with demonstrative adjectives. This does not seem to be a problem since the mother node inherits the head feature of the head noun (dmät – which is indefinite) and the phrase will be unmarked with respect to the *spec* feature. This analysis, however, misses an important characteristic of the intermediate projection, i.e. the fact that it is a definite noun phrase. This, in turn, suggests the need to introduce a separate feature in order to encode definite article just like the ones introduced in the NP based analysis. Despite such limitations, which is perhaps due to the typological differences between the two languages, i.e. German and Amharic, some of the ideas introduced by Netter can be used for the analysis of Amharic NPs as will be shown below.

Streiter (1996) proposed a similar analysis of noun phrases based on the concept of functional completeness in the context of multilingual machine translation. He claims that a three-way distinction has to be made in terms of form, function, and meaning in analyzing a language in multilingual MT. He identified determiners as functional elements that express the function of determination. In his analysis, Streiter (1996) proposed syntactic structures in which function words such as determiners head a phrasal structure, i.e. *functional-head structure* (Figure 4). Function words control the syntactic behavior of the resulting structure; as a result, the head feature of the function word percolates to the mother node. For noun phrases, the proposed structure resembles the DP analysis of Netter briefly described in the previous paragraphs, i.e. function words select

their complements in the head complement structure.

$$\left\{ \begin{array}{l} \text{role=R,} \\ \text{head=H,} \\ \text{frame=F} \end{array} \right\} \left\{ \begin{array}{l} \left\{ \begin{array}{l} \text{hpos=left,} \\ \text{role=func,} \\ \text{head}=\left\{ \text{ehead=EH} \right\} \& \text{H,} \\ \text{frame}=\left\{ \text{arg2}=\left\{ \text{head}=\left\{ \text{ehead=EH} \right\} \right\} \right\} \& \text{C} \end{array} \right\} \\ \left\{ \begin{array}{l} \text{head}=\left\{ \text{max=no} \right\} \\ \text{role=R,} \\ \text{frame=F} \end{array} \right\} \& \text{C} \end{array} \right\}$$

Figure 4 Functional-Head Structure

Another important concept is the *extended head feature*. Subsets of the head features of the function word and its complements, which determine the properties of the phrase, are grouped into extended head features and percolate up the phrasal projection. Such features are local agreement features between the function word and its complement, and features required for the selection/agreement relation of the maximal projection with a higher level predicate. In addition, features, which are related to the semantics of the phrase such as thematic roles and argument structures, are percolated from the complement to the mother node Streiter (1996:56).

Streiter (1996:59) uses a binary feature *max* (similar to *fcompl* above) in order to specify the functional completeness of a phrase. An example of such default alternation rules for nouns is given in Figure 5.

$$\left\{ \begin{array}{l} \text{head}=\left\{ \text{cat=n} \right\} \& \\ \left(\left\{ \begin{array}{l} \text{max=yes,} \\ \text{ehead}=\left\{ \begin{array}{l} \text{pform=nil,} \\ \text{type=abs,} \\ \text{wh=no,} \\ \text{neg=no,} \end{array} \right\} \end{array} \right\} ; \left\{ \text{max=no} \right\} \right) \end{array} \right\}$$

Figure 5 Default Alternation Rule for Maximal Projection (Streiter, 1996:59)

According to this rule, a bare noun may form a maximal projection in case it does not require any of the functional values, i.e. *pform=nil* (prepositions), *type=abs* (determiner), *wh=no* (wh-specifiers), *neg=no* (negation marker) otherwise it is functionally incomplete. We adopt Streiter's approach since it also takes into account the problems of multilingual machine translation system.

5. Alternative analysis

The analysis provided in this section combines the important characteristics of the approaches discussed in previous section. We start our discussion with the description of the determiners. The lexical entry for determiners such as demonstratives is given in Figure 6. The demonstrative adjective selects a nominal projection which does not already contain a specifier, i.e. *spec=no*. Furthermore, the head noun should not carry accusative case marker, i.e. *acc=no*. The *spec=no* feature among others enables to impose a unique specifier in a determiner phrase.

$$\left\{ \begin{array}{l} \text{lex=yh, lemma=yh,} \\ \text{head}=\left\{ \begin{array}{l} \text{cat=det,} \\ \text{ehead}=\left\{ \begin{array}{l} \text{type=def, ref}=\left\{ \text{deix=near} \right\}, \\ \text{agr}=\left\{ \begin{array}{l} \text{num=sing, pers=3,} \\ \text{gen=masc} \end{array} \right\} \end{array} \right\} \right\} \& \\ \left(\left\{ \begin{array}{l} \text{max=yes,} \\ \text{ehead}=\left\{ \text{pform=nil} \right\} \end{array} \right\} ; \left\{ \text{max=no} \right\} \right), \\ \text{frame}=\left\{ \text{arg1}=\left\{ \text{head}=\left\{ \begin{array}{l} \text{cat=n, spec=no,} \\ \text{acc=no, max=no} \end{array} \right\} \right\} \right\} \end{array} \right\}$$

Figure 6 Lexical Entry for Demonstrative adjectives

With respect to the different suffixes marking definiteness, nouns may be of three types: bare nouns, possessive nouns, i.e. nouns having a possessive suffix, and definite nouns, i.e. nouns with a definite article suffix. Bare nouns, both singular and plural, are underspecified with respect to definiteness feature. For example, the lexical entry for a bare singular noun is shown in Figure 7.

$$\left\{ \begin{array}{l} \text{lex=wänbär, lemma=wänbär,} \\ \text{head}=\left\{ \begin{array}{l} \text{cat=n, poss=nil,} \\ \text{ehead}=\left\{ \text{agr}=\left\{ \text{num=sing, pers=3} \right\} \right\} \right\} \& \\ \left(\left\{ \begin{array}{l} \text{max=yes,} \\ \text{ehead}=\left\{ \begin{array}{l} \text{pform=nil,} \\ \text{type}=(\text{abs; indef}), \\ \text{wh=no} \end{array} \right\} \end{array} \right\} ; \left\{ \text{max=no} \right\} \right) \end{array} \right\}$$

Figure 7 Lexical Entry for Bare Singular Nouns

The feature *poss=nil* indicates absence of possessive suffix pronouns. Nouns without determiners are interpreted as absolute or indefinite which is indicated using the disjunction *type=(abs; indef)*.

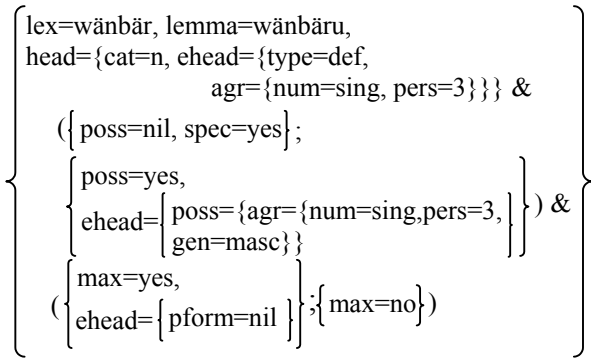


Figure 8 Lexical Entry for Definite Nouns

Nouns with the definite article suffix have the entry shown in Figure 8. We assume that the definite article is an affix and is encoded as *type=def*. The fact that the suffix *-u* may indicate the definite article and the possessive suffix for the third person singular masculine pronoun is encoded using the disjunctive head feature (*{poss=nil, spec=yes}; {poss=yes, ehead={poss=...}}*). The features representing the possessive suffix pronoun are encoded using the feature *poss={agr={num=sing, pers=3, gen=masc}}*. When the suffix is interpreted as definite article, the feature *spec=yes* is assigned. This in turn enables to specify the constraint that nouns with definite article suffix cannot be used with other determiners such as demonstrative adjectives.

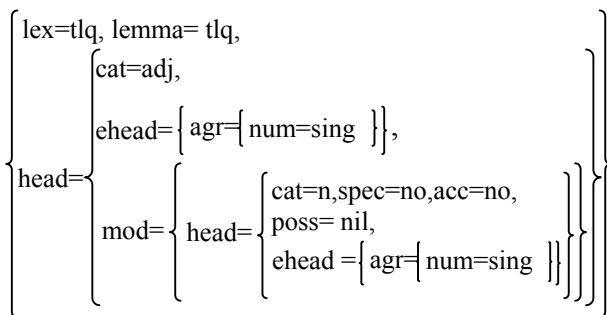


Figure 9 Lexical Entry for Indefinite Adjectives

Adjectives select nouns through the head feature *mod*. The *mod* feature imposes various restrictions on the nominal phrase. *cat=n* specifies the categorial restriction. In case of indefinite adjective (cf. Figure 9), the head noun should not have definite article suffix or possessive suffix pronouns. This is indicated by using the features *poss=nil* and *spec=no*. It should not also have the accusative case marker, i.e. *acc=no*.

For definite adjectives two minor changes were introduced (Figure 10). The definiteness of the adjective is marked with the feature, *type=def*. As with the indefinite adjective, the modified noun

should not already have a specifier. Unlike the indefinite adjectives, the definite adjective may co-occur with nouns having the possessive suffix pronoun. Therefore, the *mod* feature is underspecified with respect to the *poss* feature.

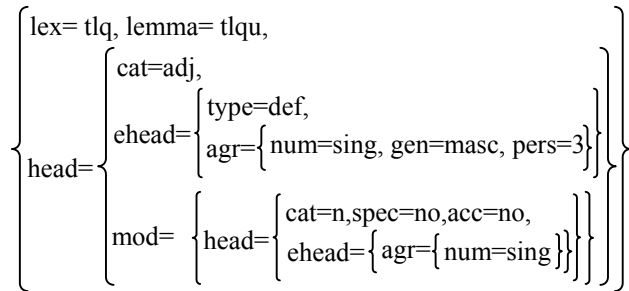


Figure 10 Lexical Entry for a Definite Adjective

The extended head feature enables to impose a co-occurrence restriction between the noun and the adjective. In addition to constraining the proper distribution of the different morphosyntactic features, the use of the extended head feature enables the proper modeling of the information flow along the phrasal projection. All the features of the daughter nodes, which determine the characteristics of the phrase, are packaged into the extended head feature and are passed over to the mother node. For example, a generic noun modified by an adjective carrying the definite article results in a definite noun phrase, the definiteness feature being contributed from the adjective. This is made possible since definiteness is part of extended head feature and both the adjective and the head noun unify their extended head features (Figure 10).

In general, the above analysis also provides an account of certain characteristics of Amharic noun phrases which pose problems to the transformation-based treatment suggested by Girma (1994:108) (cf. Figure 1). His analysis is based on the assumption that ‘an NP is complement of a single D’. Therefore, his analysis falls short of an explanation of noun phrases containing a cluster of adjuncts with multiple Ds as shown in (19). In (19), the noun *däbtär* is modified by two attributive adjectives both of which contain the definite article.

19. tlq-u tqur-u däbtär
big-DEF black-DEF notebook
‘the big black notebook’

On the other hand, the analysis provided in this section easily extends to (19) without the need to change the grammar.

6. Application to MT

The importance of the analysis provided in the previous section for machine translation has been discussed at length in Streiter (1996) using the CAT2 MT system, which is a transfer-based machine translation system. The translation process is divided into smaller subtasks, which usually correspond to the general steps in the analysis of natural language, i.e. morphology, syntax, and semantics. Transfer into the target language is carried out based on a well-defined canonical structure, i.e. the Interface Structure (IS). The interface structure abstracts away from the idiosyncratic properties of the source and target language, and maps several monolingual structures onto a few sets of interface structures. This in turn simplifies transfer, as there will be fewer structures

to be transferred.

Identification and description of similarities and differences between the languages involved in the translation process constitutes an important task in the design of the interface structure. The form and use of functional words are typical grammatical aspects in which languages differ radically (Streiter, 1996:25-26). This seems to be the case for Amharic and English too. Unlike in English, in Amharic most function words such as determiners, prepositions, and conjunctions are generally bound morphemes. For example, in Amharic, the definite article is a suffix morpheme whereas in English it is an independent word. Then it is more attractive to treat it using a set of semantic features in IS leaving the actual surface realization to the monolingual component. As a first step towards this goal, the definite article is encoded as a feature in the extended head feature of the constituent to which it is attached, as shown in (Figure 10). The extended head feature will be transferred to the target language. The target language component

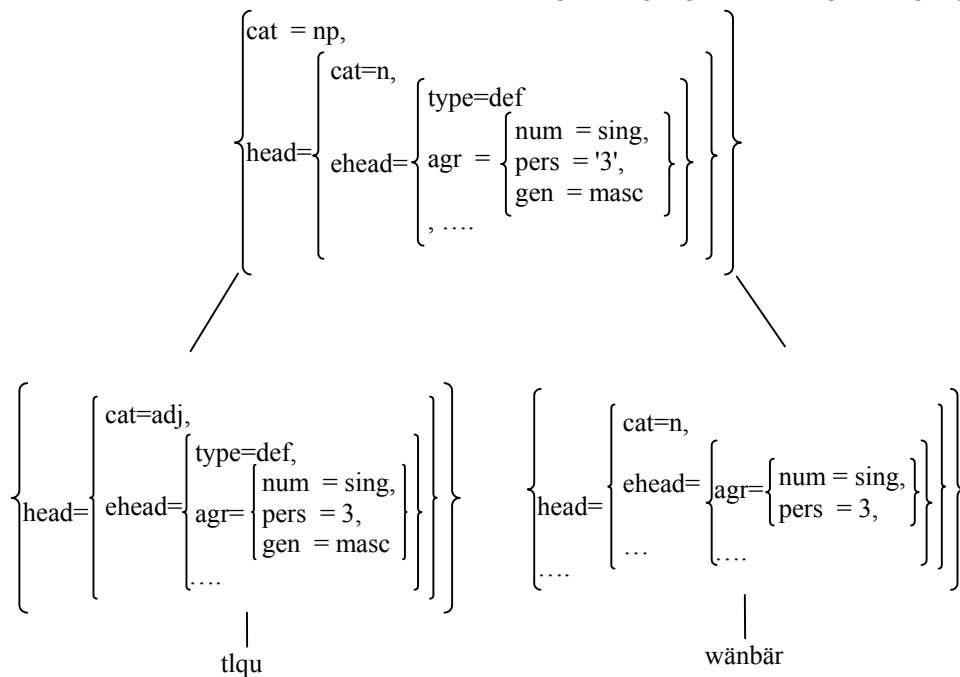


Figure 10 An Example Definite Noun Phrase

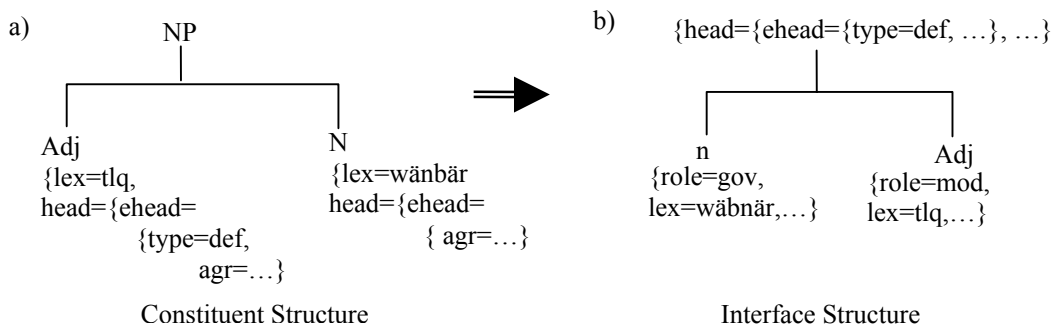


Figure 10 An Example Interface Structure Representation

will then decided how the features used to encode definiteness are realized on the surface in the target language.

7. Summary

In this paper, we showed that lexical-based approaches explain the characteristics of Amharic noun phrases better than the transformational-based treatments suggested in Amharic grammar literature. Furthermore, the fact that Amharic definite article is a bound morpheme makes it difficult the direct application of the different strategies suggested in the literature, i.e. NP analysis and DP analysis, for the analysis of Amharic noun phrases. As a result, we combined some features of NP and DP analysis. Following Wintner (2000), we treated Amharic definite article as an affix and represented it as feature of a head noun. We adopted the mechanism proposed by Netter (1996) for the specification of intermediate projection, and the general syntactic analysis strategy suggested by Streiter (1996). This provided us with the necessary mechanisms to account for the phenomena in Amharic noun phrases. Our analysis is also in line with the mechanism suggested by Streiter (1996) for simplifying the design of interface structure for a transfer-based machine translation system.

References

- Abiyot, Bayou, (2000), Developing automatic word parser for Amharic verbs and their derivation, MSc Thesis, Addis Ababa University, Addis Ababa.
- Abney, Steven, (1987), The English Noun Phrase in Its Sentential Aspect, Ph.D. Dissertation, MIT, Cambridge.
- Atelach, Alemu, (2002), Automatic Sentence Parsing for Amharic Text: An Experiment Using Probabilistic Context Free Grammars, MSc. Thesis, Addis Ababa University, Addis Ababa.
- Baye, Yimam, (1986), *yāamarña säwäsäw*, Addis Ababa, E.M.P.D.A.
- Bernstein, Judy B., (2001), 'The DP Hypothesis: Identifying Clausal Properties in the Nominal Domain', in: Mark Baltin and Chris Collins, editors, *The Handbook of Contemporary Syntactic Theory*, Blackwell Publishers, Malden, 536-561.
- Bresnan, Joan (ed.), (1982), *The Mental Representation of Grammatical Relations*, Cambridge, Mass., The MIT Press.
- Girma, Halefom, (1994), The Syntax of Functional Categories: a Study of Amharic, PhD Dissertation, Montreal, Université du Que'bec.
- Haller, Johann, (1993), 'CAT2: Vom Forschungssystem zum präindustriellen Prototyp', in: Horst P. Pütz and Johann Haller, editors, *Sprachtechnologie: Methoden, Werkzeuge, Perspektiven*, 282-303, Hildesheim. GLDV, Georg Olms AG, URL: <http://www.iai.uni-sb.de/cat2/docs.html>.
- Leslau, Wolf, (1995), *Reference Grammar of Amharic*, Otto Harrassowitz, Wiesbaden.
- Mesfin, Getachew, (2001), Automatic part-of-speech tagging for Amharic language: An experiment using stochastic HMM, MSc Thesis, Addis Ababa University, Addis Ababa.
- Nega, Alemayehu, (1999), Development of Stemming Algorithm for Amharic Text Retrieval, PhD Dissertation, University of Sheffield.
- Netter, Klaus, (1996), Functional Categories in an HPSG for German, PhD Dissertation, University of Saarland, Saarbrücken.
- Pollard, Carl; Ivan Sag, (1994), *Head-Driven Phrase Structure Grammar; Studies in Contemporary Linguistics*, Chicago, London, The University of Chicago Press.
- Sisay, Fissaha; Johann Haller, (2003a), 'Amharic verb lexicon in the context of machine translation', in: *Traitement Automatique des Langues Naturelles*, Vol. 2, 183-192.
- Sisay, Fissaha; Johann Haller, (2003b), 'Application of corpus-based techniques to Amharic texts', in: *Machine Translation for Semitic languages, MT Summit IX Workshop*, New Orleans.
- Streiter, Oliver, (1996), Linguistic modelling for Multilingual Machine Translation, PhD Dissertation, University of Saarland, Shaker Verlag.
- Wintner, Shuly, (2000), 'Definiteness in the Hebrew noun phrase', in: *Journal of Linguistics*, 36, 319-363.
- Zwicky, Arnold M., (1985), 'Clitics and particles', *Language*, 61(2), 283-305.