ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

SCHOOL OF INFORMATION STUDIES FOR AFRICA

DEVELOPING ENGLISH-TIGRINYA MACHINE-READABLE BILINGUAL DICTIONARY: AN INPUT FOR MACHINE TRANSLATION

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION SCIENCE

BY

MULU GEBREEGZIABHER

JUNE 2001

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

SCHOOL OF INFORMATION STUDIES FOR AFRICA

DEVELOPING ENGLISH-TIGRINYA MACHINE-READABLE BILINGUAL DICTIONARY: AN INPUT FOR MACHINE TRANSLATION

BY

Mulu Gebreegziabher

Name and Signature of Members of the Examining Board

to Getachew Jemaneh, Chairman, Examining Board

r. Gebremedhin Simon, Advisor

to Workshet Lamenew, Advisor

r. Kamal Bechkoum, External Examiner

ACKNOWLEDGMENT

I am very much indebted to my thesis advisors Dr. Gebremedhin Simon and Ato Workshet Lamenew, whose valuable advice, discussion and critical supervision enabled me to write this thesis.

Special thanks are also due to Ato Tesfaye Biru for his great encouragement in choosing the thesis's topic. I am also thankful to Ato Sisay Fissaha, who helped me to narrow my research topic and to Ato Weldu Mickael for lending me books and for his very helpful suggestions and comments.

Finally, I would like to express my deep indebtedness to my fiancée W/t Netsanet Kebede and to all friends of mine, who in one way or another helped me in so many things.

TABLE OF CONTENTS

CONTENTS **PAGES**

DECLARATION ii

ACKNOWLEDGMENT iii

TABLE OF CONTENTS iv

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ALGORITHMS ix

LIST OF ABBREVIATIONS AND SYMBOLS USED x

ABSTRACT xii

CHAPTER ONE 1

INTRODUCTION 1

1.1 Background 1

1.2 Statement of the problem 10

1.3 Objective 14

1.3.1 General objective 14

1.3.2 Specific objectives 14

1.4 Scope of the study 14

1.5 Significance of the study 15

1.6 Methodology 16

1.6.1 Literature review 16
1.6.2 Development tools 16
1.6.3 Testing technique 17
1.7 Organization of the thesis 17
CHAPTER TWO 19
MACHINE TRANSLATION 19
2.1 Introduction 19
2.2 Rule-based approaches 20
2.2.1 Transfer systems 20
2.2.2 Interlingual systems 22
2.3 Corpus-based approaches 23
2.3.1 Statistical methods 23
2.3.2 Example-based approaches 24
2.4 Contents of machine-readable dictionaries 25
2.4.1 Morphology 26
2.4.2 Syntax 26
2.4.3 Semantic 27
2.4.4 Concepts 27
CHAPTER THREE 29
ENGLISH AND TIGRINYA NOUNS: A CONTRASTIVE ANALYSIS 29
3.1 Introduction 29
3.2 Syllable 29
3.3 Noun formation 33

1.6.1 Literature review 16
1.6.2 Development tools 16
1.6.3 Testing technique 17
1.7 Organization of the thesis 17
CHAPTER TWO 19
MACHINE TRANSLATION 19
2.1 Introduction 19
2.2 Rule-based approaches 20
2.2.1 Transfer systems 20
2.2.2 Interlingual systems 22
2.3 Corpus-based approaches 23
2.3.1 Statistical methods 23
2.3.2 Example-based approaches 24
2.4 Contents of machine-readable dictionaries 25
2.4.1 Morphology 26
2.4.2 Syntax 26
2.4.3 Semantic 27
2.4.4 Concepts 27
CHAPTER THREE 29
ENGLISH AND TIGRINYA NOUNS: A CONTRASTIVE ANALYSIS 29
3.1 Introduction 29
3.2 Syllable 29
3.3 Noun formation 33

3.4 Gender 38
3.5 Number (singular or plural forms) 39
3.6 Possessive pronominals 47
3.7 Derivational nouns 51
3.8 Direct/Indirect objects 55
3.9 Definite/indefinite articles 56
CHAPTER FOUR 58
DESIGN OF THE PROTOTYPE SYSTEM 58
4.1 Introduction 58
4.2 The structure of the English-Tigrinya Bilingual Dictionary 59
4.2.1 Headword information 61
4.2.2 Grammatical information 67
4.2.3 Semantic information 69
4.2.4 Management information 69
4.3 Design of the English-Tigrinya Bilingual Dictionary 70
4.4 Algorithms for translating nouns and the experiment 71
4.4.1 Affixes to nouns 72
4.4.2 Developing algorithms for affixes 76
4.4.3 The experiment and the test result 83
CHAPTER FIVE 88
CONCLUSIONS AND RECOMMENDATIONS 88
5.1 Introduction 88
5.2 Conclusion 89

5.3 Recommendations 91

Bibliography 93

APPENDICES 98

APPENDIX 1: Database record structure and sample data 98

APPENDIX 2: Translated sample noun forms of English and Tigrinya languages102

LIST OF TABLES

Table 3.1: Simple nominal stems derived from triradicals 36

Table 4.1: Structure of English-Tigrinya Bilingual Dictionary Record 60

Table 4.2: The left-adjacency attributes 63

Table 4.3: The right-adjacency attributes 64

Table 4.4: The grammatical information 67

Table 4.5: Attributes and data types of the English-Tigrinya Bilingual Dictionary70

Table 4.6: Noun-endings for regular English nouns 73

Table 4.7: Affix orders to Tigrinya nouns 74

Table 4.8: The combination of noun prefixes in Tigrinya 75

Table 4.9: the combination of noun suffixes in Tigrinya 75

LIST OF FIGURES

Figure 2.1: Schematic of document translation utilizing the transfer method 21

Figure 3.1: The structure of Tigrinya syllable..... 31

Figure 3.2: Three tiers for Tigrinya stem..... 34

Figure 4.1: An overview diagram for translating English and Tigrinya nouns..... 71

LIST OF ALGORITHMS

Figure 4.2: The algorithm for processing English and Tigrinya plural noun forms76

Figure 4.3: The algorithm for processing indefinite article for English nouns 78

Figure 4.4: The algorithm for processing definite article for English nouns 79

Figure 4.5: The algorithm for processing definite article for Tigrinya nouns 80

Figure 4.6: An algorithm for Tigrinya possessive pronominals 81

LIST OF ABBREVIATIONS AND SYMBOLS USED

1c.	=	1st person
2f.	=	2nd person female
2m.	=	2nd person male
3f.	=	3rd person female
3m.	=	3rd person male
CV	=	Consonant Vowel
C^w	=	labialized consonant
f.	=	feminine (female)
f.pl.	=	female plural
f.s.	=	female singular
m.	=	masculine (male)
m.pl.	=	male plural
m.s.	=	male singular
pl.	=	plural
s.	=	singular
SR	=	Surface Representation
sth	=	something
UR	=	Underlying Representation

VOWELS

ɨ	=	high central unrounded vowel
ä	=	mid central unrounded vowel

CONSONANTS

ʕ	=	voiced pharyngeal fricative
č	=	voiceless palatal affricate
ħ	=	voiceless pharyngeal fricative
č‘	=	voiceless glottalized palatal affricate
?	=	voiceless laryngeal stop
k’	=	voiceless glottalized velar stop
p’	=	voiceless glottalized bilabial stop
š	=	voiceless palatal fricative
s’	=	voiceless glottalized dental fricative
t’	=	voiceless glottalized dental stop
x	=	voiceless velar fricative
x’	=	voiceless glottalized velar fricative
ň	=	voiced nasal palatal

ABSTRACT

The cultural, political, social, scientific and technological information of the advanced world including our country is mostly published in English language. Although English language is widely used in the country, the language barrier makes it difficult to use such information by the majority of the population.

Translation is the best solution to overcome such language barrier. To translate such information to local languages using the human skill is on the one hand the oldest and the most accurate method on the other hand it is expensive and time consuming task. Nowadays, it is possible to translate a digitized text from one language into another using machine translation systems. As a result, cost is minimized and speed is maximized.

However, to develop a machine translation system is a complex task. To reduce the complexity, the other alternative is to develop translation modules and integrate them rather than to develop self-contained and independent systems. Machine-readable dictionary is the main component or module of the machine translation system.

This study therefore focuses on developing English-Tigrinya Bilingual Dictionary that can be used as input for machine translation system. The stated dictionary contains a lexicon for English and Tigrinya nouns. The noun forms and functions of both languages have been discussed and an attempt is made to compare the major similarities and differences.

Based on the noun forms and functions of both languages, the bilingual dictionary is designed and the database for storing the lexicon is constructed. Next, a list of hundred English nouns and hundred Tigrinya nouns are entered to the database. Finally, an algorithm is developed to translate English noun forms to Tigrinya noun forms.

CHAPTER ONE

INTRODUCTION

1.1 Background

Tigrinya is a Semitic language spoken in the Tigray Regional State, parts of Wollo and Gonder, and in the Eritrean Highlands (Tesfai 1993, Oriolo et al 1997). Together with Geez and Tigre it forms a group with in Semitic known as North Ethiosemitic (Tesfai 1993). Tigrinya is the most widely spoken Semitic language in the world next to Amharic (Girmay 1983, Bender, et al 1976). As Amharic is the third widely spoken Semitic language, Tigrinya is the fourth after Arabic, Hebrew, and Amharic (Oriolo et al 1997).

In the thirteenth century, documents concerning Tigrinya literature were found in Šimezana (in Akkeleguzay - Southern Eritrea). But at that time, the language is said to have only little literature (Tesfai 1993). Tigrinya developed, in its written form, only after the beginning of the nineteenth century (Oriolo et al 1997).

Apparently, Tigrinya is serving as a working language of the Tigray Regional State. Together with English, Tigrinya is serving as a teaching learning language in the educational system of Tigray.

English is widely used and is the first foreign language of the country. English is used as a medium of instruction in high schools and universities. It is widely used in workshops, seminars, conferences, etc. organized by governmental and non-governmental organizations. There are also mass media that use English. There

are English programs transmitted by the Ethiopian radio and television. Private and governmental English newspapers, newsletters, magazines, journals, pamphlets, etc. are also widely available.

Although English language is widely used in the country, with the exception of educated people and foreigners, the majority of the population does not comprehend this language. There are also many journals, books, newspapers, etc. published abroad in foreign languages that might be very useful to the country (Kinfe 1972 E.C.). For example, scientific and technological information is mostly published in English. However, the language barrier makes it difficult to use such information by the mass. As Solomon (1969) points out, there is no reason for every Ethiopian to know foreign language to have the world's great books available to him.

However, in this world of globalization, one may ask the need for having literature in mother tongue rather than using the readily available literature in English or other foreign languages. Studies have been conducted for justifying the need for publishing literatures using the local languages.

A paper presented by Kinfe (1972 E.C.) at a translation seminar points out the disadvantages of using English or other foreign language literature by the local population. Some of these limitations are:

- Foreigners do researches about the history, literature, and culture of Ethiopian nations and nationalities. The educated Ethiopians not only accept the views of the foreigners but also teach students as it is presented by the foreigners with out further analyzing the research.

- Ethiopian researchers produce their researches by using either the foreign language, such as English, French, Italian, etc. or Amharic, which is the national language of the country. Thus, the local people are facing difficulties in understanding and using not only literatures written in foreign languages, but also literatures written in Amharic for those whose mother tongue is other than Amharic.

At this same seminar, another paper presented by the Curriculum Department (1972 E.C.) has also put the negative impact of literature written in foreign languages on the educational system of the country. Thus,

- Students cannot understand the handbooks written in English.
- The students face difficulties in understanding what is said in the class as much of the presentation is in English.
- Students are obliged to know about the western countries rather than their own country or environment.
- The literatures reflect the cultural, political, and socioeconomic conditions of the advanced societies rather than of the Ethiopians.
- Scientific and technological terms are known or are used by students and teachers but not by the mass. Accordingly, every body uses the terms by giving their own meanings.

On the other hand, there are many advantages to have literatures in mother tongue. Wedekind (1972 E.C.) indicates the advantages:

- The mother tongue is the core identity of the people's culture.
- To use both the mother tongue and other languages at the same time is an advantage and is just like becoming a bicultural or polyglot.
- The literature is more convenient to the readers so that they can benefit from it. However, if the literature is in foreign languages, the readers switch off and ignore the publication.
- The literatures should express the specific bicultural bilingual conditions of the indigenous group, with absolute respect for its cultural uniqueness and human identity.
- It is desirable to provide literacy and literature in every body's mother tongue.
- Illiteracy may decrease.
- The mother tongue is the principal vehicle for acquiring both linguistic and non-linguistic knowledge.
- Fewer dropouts from schools, fewer failures in exams, increase of total number of students and greater fluency in the national language.
- Greater number of community projects from bilingual schools, greater reading comprehension and faster learning of the national language; a larger proportion of adult literacy.
- In the field of education and social advance, a secure passage toward integration will be opened only through the use of the native language.

So, how can the local people benefit from useful literature written in foreign languages? How can the mentioned disadvantages on using literatures written in foreign languages be minimized? Dictionaries can be useful tools for local peoples in understanding and using literature written in foreign languages. Besides, they can facilitate the translation of literature of foreign origin to local languages. The currently available Tigrinya dictionaries however have limitations in many respects. Before discussing the limitations, the available bilingual Tigrinya dictionaries are discussed first.

The introductory of the dictionary 'mäzgäbä k'alat Tɨgriňňa-bɨTɨgriňňa' published by the Academy of Ethiopian Languages (1989 E.C.) points out that there are printed bilingual dictionaries such as Tigrinya-Italian by Basano in 1918, Tigrinya-Amharic by Yohanes G/egziabher in 1949, and Tigrinya-French by Coulbeaux and Schreiber in 1915.

There are also dictionaries that target English and Tigrinya readers. The Eritrean People's Liberation Front (EPLF), for example, prepared Tigrinya-English dictionary in 1986. The primary purpose of this dictionary is to help those Tigrinya speakers who know English well to read Tigrinya literature particularly to literature of the EPLF. As there is minimal Tigrinya literature, the secondary purpose of this dictionary is to promote and encourage the literate (EPLF 1986).

The other is a multilingual dictionary of English-Tigrinya-Italian prepared by Oriolo et al in 1997. The main purpose of the dictionary is to provide essential practice for developing the learner's ability to communicate effectively in the mentioned

languages. It also provides sentence constructions using the three languages that can help tourists (Oriolo et. al 1997).

There are also others such as the 'Self-Study Tigrinya for Foreigners and English for Tigrinya Speakers' by Mesfin Ghebrehiwet in 1996; 'Concise English-Tigrinya Dictionary' by Tuquabo Aressi in 1987; 'Tigrinya-English Dictionary' by Research and Information Center of Eritrea in 1986; and 'English-Tigrinya Dictionary' by Uqbamicael Habtemariam in 1993.

The above-mentioned dictionaries have limitations. A symposium on dictionary preparation in Ethiopia was held in 1983 by the Academy of Ethiopian Languages (1975 E.C.). A study on Tigrinya dictionaries, which was presented in the symposium, indicates that Tigrinya dictionaries prepared until 1983:

- were prepared using European languages except the dictionary prepared by Yohanes G/egziabher
- some of the dictionaries use Geez for the head words
- except Yohanes G/egziabher, the authors do not know the Tigrinya language however they managed to prepare the dictionaries with the help of translators
- were mostly prepared by Italian Missionaries during the Italian occupation of Ethiopia. As a result, the preparation of dictionaries was not a continuing process.
- have limited words from 2,000 to 10,000

- their purpose is to help foreigners communicate with Tigrinya speakers while they are in Ethiopia

Dictionaries prepared since 1983 on the other hand target most literate Tigrinya speakers (who are exposed to read Amharic or foreign literature) as well as foreigners (EPLF 1986, Oriolo et al 1997). They are helpful for tutoring purpose to foreigners as well as to Tigrinya language speakers. Therefore, these dictionaries are supposed to encourage and expose foreigners to read Tigrinya literature and Tigrinya speakers to read foreign literature.

The electronic version of the conventional printed dictionaries is also called digital online dictionary. This online dictionary is designed as online reference books for use by humans. Typically, they contain the exact textual contents of a conventional printed dictionary stored on magnetic disk or CD-ROM. Usage is analogous to a conventional dictionary, except that the electronic version may have hypertext-link features for more rapid and convenient browsing (Kahaner 1992).

The printed or online conventional dictionaries are only used by humans for reference purposes. As far as the student researcher is concerned, no other printed dictionaries other than the mentioned ones are available. Besides, there are no Tigrinya online dictionaries available at all.

The other types of dictionaries are machine-readable dictionaries. Computers use such dictionaries for different purposes such as machine translation. The machine-readable dictionary used for the purpose of machine translation, here after MT, must be different from conventional dictionaries described earlier. That is, the machine-readable dictionary can use the possibilities of computer in view of cross-reference,

morphological analysis of inflections, synonyms, linguistic concepts and additional grammatical information. As a result, the machine-readable dictionaries can be useful for analyzing the source language and for generating the target language (HEISOFT 2000).

One of such dictionaries used for MT purpose is the Electronic Dictionary Research (EDR). The EDR project was established in 1986 as a consortium in collaboration with eight member corporations with a nine-year charter to develop a large-scale practical machine-readable dictionary system that could be used in support of a variety of natural language tasks, such as translation between English and Japanese, natural language understanding and generation, speech processing, and so forth.

The EDR's primary goal is to capture all the information a computer requires for a thorough understanding of natural language. This includes the meanings (concepts) of words; the knowledge needed for the computer to understand the concepts; and the information needed to support morphological analysis and generation, syntactic analysis and generation, and semantic analysis.

EDR has adopted a pragmatic, example-driven approach to the construction of the dictionary system. As Kahaner (1992) points out, some early EDR work was done with Prolog, although now most of the programming is in C. While stronger theoretical work and better tools are both necessary and desirable, there is no substitute for breadth of coverage and the hard work necessary to achieve it.

The breadth of the EDR can be expressed at least by looking at the set of component dictionaries, which serve distinct roles though collectively form, the EDR

product. These include word dictionaries for English and Japanese having 200,000 words each, the Concept Dictionary with 400,000 concepts, co-occurrence dictionaries for English and Japanese having 300,000 words each and two bilingual dictionaries: English to Japanese and Japanese to English having 300,000 words each (Kahaner 1992).

Researches from inside and outside Japan indicate that Japanese has greater commitment to MT (Kahaner 1992). In Japan, there are MT systems having dictionaries ranging in size from 50,000 to 800,000 entries but none are as rich or as detailed as the EDR dictionary. EDR has accomplished this breadth and this is an essentially unique achievement. Kahaner (1992) further elaborates that the comparison between the US and Japan in terms of MT and related technologies shows that Japan is ahead of the US in several important ways. These include commercial use of MT, acceptance of MT by users and above all the development of knowledge sources such as dictionaries.

The EDR staff also encourages international cooperation in the use and further development of the dictionaries. As Kahaner (1992) indicates, the same conditions regarding the use of the EDR dictionaries are being applied to all users of domestic (Japanese) or overseas. Especially for academic users such as universities and public research institutions, special measures are being considered including very low prices.

For developing the prototype English-Tigrinya machine-readable bilingual dictionary, the EDR system will be used. The current study attempts to build a prototype English-Tigrinya machine-readable bilingual dictionary that can be used as input for

MT. Hence, it can be considered as a promising step for the development of Tigrinya in the computing area.

The part of speech included in this research is however common nouns of both languages. The three major lexical categories in Tigrinya are nouns, verbs, and adjectives (Tesfai 1993). The adverbial functions are accomplished by other categories. For example, the combination of the preposition /bɨ-/ 'by' and the adjective /qɨlt'uf/ 'quick' give the adverb /bɨqɨlt'uf/ 'quickly'. Among the major lexical categories in Tigrinya, the morphological analysis and generation of verbs has been done by Markus Walther in 1998. The result of the computational model is being posted on the Internet. Although, its target is not for building machine-readable dictionary, it computes all the verbal forms in Tigrinya language. It performs the parsing and generation of verbal words. For the reasons that the output is still not available in its complete form and not to do duplication, the noun lexical category is chosen for this research.

1.2 Statement of the problem

Translation is the best solution to overcome the language barriers. The white paper of the Transparent Language Inc. (2000) indicates that translation demands a high degree human skill in the areas of language comprehension, recognition and memory for at least two languages. Human translators use a variety of thought processes, skills and resources to interpret the meaning of a sentence and communicate the meaning of that sentence in a different language. Pang (1998) indicates that human translation service is the oldest and the most accurate method. Pang (1998) adds, the human translator can understand the purpose of the

document and can realize the real situation and translates according the cultural context of the country by making sure that there is no lose in meaning, relevance and context of the original source text.

Although translation using human skill is preferred to produce accurate translation, it is globally accepted as an expensive and time-consuming task (Transparent Language 2000). The problem of human translation services is not different in the case of our country. The Adult Education Department (1975 E.C.) has identified the following as the major problems encountered as a result of human translation services:

- Lack of skilled manpower.
- Finance.
- Lack of dictionaries published using the local languages.
- Lack of subject dictionaries.
- Even some of the dictionaries published in local languages are not updated regularly.
- The word forms of the local dictionaries are not complete so that translators are obliged to use the English terms.
- In some cases, It is difficult to get equivalent terms in local languages for some foreign terms and there is unwillingness to use such untranslatable foreign terms.

- There is no responsible body to translate new political, economical, social, cultural and scientific and technological terms so that inconsistency is introduced in using or translating such terms.

- Translated books may not be accepted simply because the dialects of the different regions may not be included in the translated books.

Some of the above mentioned problems are solved using MT systems. Using MT systems, cost is minimized and speed is maximized. These two are especially the current demands of the business community. The competitive business community communicates world wide on daily bases using the Internet technology. Therefore, they need an urgent inter-language communication tool that can at least provide a draft-quality translation to determine or judge the relevance of the documents. MT is also a useful tool for human translators and has been proved that it increases their productivity by 30% or more. (Transparent Language Inc. 2000)

Maegaard (2000) points out that ten years ago MT has been part of linguistic services, which may be as diverse as E-mail across nations. Using the Internet technology, we can get access to digitized texts of diversified languages. The MT plays a major role to translate these digitized texts available on the Internet from one language to another. According to Maegaard (2000), it has been not only a tool for translating large Internet texts and traditional documents but also it has been used as a portable tool including speech translation for travelers, soldiers, etc.

Apparently, the situation is fast changing due to the ongoing commercial growth and the influence of new research. MT is becoming a tool for translation by freelancers and smaller organizations. Cheap translation assistants are making their way to

market to help small companies and individuals write foreign letters, E-mail, and business reports. Furthermore, given the explosion of information in the world, the demand for translation is booming, and the only possible answer to this demand is MT (Maegaard 2000).

Hutchins (1994) indicates that the first MT began in 1949. Since then, the western countries and Japan have conducted a series of researches. At their early stage, MT projects were brought to an end. MT projects fail, according to Hutchins (1994), not only because they are very complex and expensive (that is, the initial cost is high) but also they fail to meet their promises. However, through time research on machine translation regain by developing translation modules to be integrated rather than the design of systems to be self-contained and independent.

Machine-readable dictionary is the main component or module of MT software. Unless the machine-readable dictionary is designed carefully, the output of the MT might be disappointing or totally meaningless. You can imagine if a single word of a sentence to be translated is missing in the dictionary. The machine finds it difficult to understand the whole sentence. Even if all words of a sentence are in the dictionary, the MT may interpret it wrongly due to the ambiguity of words or phrases (HEISOFT 2000).

Therefore, in view of the above stated ideas, the development of English-Tigrinya machine-readable dictionary would help for machine translation. As to my knowledge, there is no such research conducted. This study focuses on investigating the possibility of developing English-Tigrinya machine-readable bilingual dictionary.

1.3 Objective

1.3.1 General objective

The general objective of the study is to explore the possibility of developing an English-Tigrinya machine-readable bilingual dictionary that could be used as an input for machine translation. The purpose of this study is to test the EDR system for building English-Tigrinya bilingual dictionary.

1.3.2 Specific objectives

Specifically, this study attempts:

- To study the forms of Tigrinya nouns.
- To study the functions of Tigrinya nouns.
- To study the similarities and differences between English and Tigrinya nouns.
- To develop an algorithm for the English-Tigrinya Bilingual Dictionary.
- To develop the prototype English-Tigrinya machine-readable bilingual dictionary using the tools, techniques and methodologies used to develop the EDR bilingual dictionary.

1.4 Scope of the study

The scope is limited to build a prototype English-Tigrinya machine-readable bilingual dictionary since it is not possible to build a fully-fledged bilingual machine-readable dictionary within the time allotted for the study. Besides, the bilingual dictionary

contains the forms and functions of English and Tigrinya noun lexical category. In this study, other parts of speech such as verbs, adjectives, adverbs, prepositions, etc are excluded. The study also excludes Proper nouns such as names of places and people, cardinal numbers such as one, two, etc., ordinal numbers such as first, second, etc., and classifiers such as piece, amount, etc.

1.5 Significance of the study

The current study has a direct or an indirect impact for the development of different natural language processing in the computing area of Tigrinya language. The dictionary:

- is used as an input to machine translation from English to Tigrinya and vise versa
- can be used as a model for building verbs, adverbs, pronouns, phrases, etc. of Tigrinya language
- can be used as a model for building subject bilingual dictionaries
- can be applied with modifications for word processors or spell checkers
- can be used by Tigrinya speakers and foreigners to know the equivalent of English and Tigrinya nouns
- can be applied for automatic abstracting systems that need dictionaries to understand contents of a document
- Can fill the gap in Tigrinya language computing

- Can be used as the basis to develop bilingual dictionaries for other languages of the country

1.6 Methodology

In order to achieve the final goal, the following methods have been followed:

1.6.1 Literature review

To have a clear picture for the development of the English-Tigrinya machine-readable bilingual dictionary, data must be collected. The kind of linguistics data needed for building machine-readable dictionary in general is more detailed than the conventional dictionaries.

Accordingly, data that show the morphology, semantic and grammatical properties of Tigrinya nouns have been studied from sources on Tigrinya language. Information that contribute to the comparative study of the corresponding English and Tigrinya languages have been reviewed and analyzed in order to identify major differences and similarities between the two languages. Furthermore, literatures that discuss the EDR system have been reviewed.

1.6.2 Development tools

Based on the result of the analysis, a prototype English-Tigrinya machine-readable bilingual dictionary is developed using the Microsoft Visual C++ programming language. The Microsoft Visual C++ is chosen in that it is the window version of C programming language, which is used for the construction of the EDR machine-readable dictionaries. As one aim of this study is to test the EDR system, it has to

employ the tools and techniques used by EDR. The Microsoft Visual C++ is also a powerful tool for development, extension and modification of programs. The programs can be easily integrated with Microsoft Access that is used for storing the data for the prototype and it incorporates user interfaces and is easy to add menus and other features.

1.6.3 Testing technique

Using the techniques and tools employed by EDR system, the possibility for construction of the bilingual English-Tigrinya dictionary is explored and tested. The system is tested using selected English and Tigrinya sample noun forms. Paper-based dictionaries are used for constructing the different noun forms of both languages. The corresponding noun forms of the two languages are tested again and again until the desired output is achieved. Based on the exploration or test result, a decision is made among the choices of using or customizing the EDR system or building a new system.

1.7 Organization of the thesis

The thesis is organized in five chapters. The first chapter discusses the background information, statement of the problem, objective, scope, significance and methodology of the study. Chapter two discusses the approaches that are followed by different machine translation systems and the contents of bilingual dictionaries, which are the basic components of the machine translation systems. The third chapter reviews the forms and functions of English and Tigrinya nouns that are used in the construction of the prototype English-Tigrinya Bilingual Dictionary. Chapter

four discusses the structure, content, algorithms, coding, and discussions of the prototype English-Tigrinya Bilingual Dictionary. Finally, chapter five discusses the conclusion and recommendation of the study. Bibliographies and appendices are also attached at the end of the thesis.

CHAPTER TWO

MACHINE TRANSLATION

2.1 Introduction

The beginning of machine translation, according to Hutchins (1994), is in 1949 with the memorandum from Warren Weaver. Since then, MT has passed a series of stages. The second stage began in 1954 when a demonstration of a simple translation system from Russian to English was made. At this stage the basic processing method for MT, as is reported by Transparent Language Inc. (2000), was the 'direct method'. The direct method translates word by word without interpreting sentence structure. The result is a translation that is in the word order of the source text, which can result in a significant loss of meaning and readability and the comprehension and translation quality is often unacceptable. This stage, as Hutchins (1994) points out, was brought to an end in the United States in 1966 and had strongly affected MT research elsewhere. The major reason was a report made by ALPAC that highlighted the failure of MT research to meet its promises. The third stage lasted until 1975 when MT was ignored and is called the 'quiet era'. The next stage that indicates the beginning of a 'new era' lasted until the end of 1980s basically with the development of new rule-based methods followed by the emergence of corpus-based approaches since 1989.

2.2 Rule-based approaches

As Hutchins (1994) indicates, the dominant MT research from 1984 until the end of the 1980s was the approach based on essentially linguistic rules of various kinds: rules for syntactic analysis, lexical rules, rules for lexical transfer, rules for syntactic generation, rules for morphology, etc. The dominant MT systems of that time were the 'transfer' based MT systems and various 'interlingual' systems appeared in the later 1980s. Most of these systems were basically linguistics-oriented, but some adopted knowledge-based approaches, making use of non-linguistic information relevant to the domains of texts to be translated. However, these newer knowledge-based systems are rule-based and continued until almost the end of the decade.

The common feature of all pure rule-based MT systems is that they involve many costs in formulating rules, confirms Chen (1995). It is also easy to introduce inconsistencies in a rule-based system, when the number of rules increases. As Maegaard (2000) clearly points out, commercial MT systems that contain between a quarter and a half million words can have up to thousand grammar rules for each of the languages. Therefore, more rules lead to more sophistication and more complexity, and may in the end develop into systems that are quite difficult to maintain.

2.2.1 Transfer systems

The rule-based transfer system is basically linguistics-based system and applies such like syntax-oriented approaches, explains Hutchins (1994). The transfer systems have three stages of analysis, transfer and synthesis and with the

processes of analysis and generation passing through series of distinct levels (morphology, syntax, and semantics). Figure 2.1 is an example of a schematic representation of a transfer-based MT system.

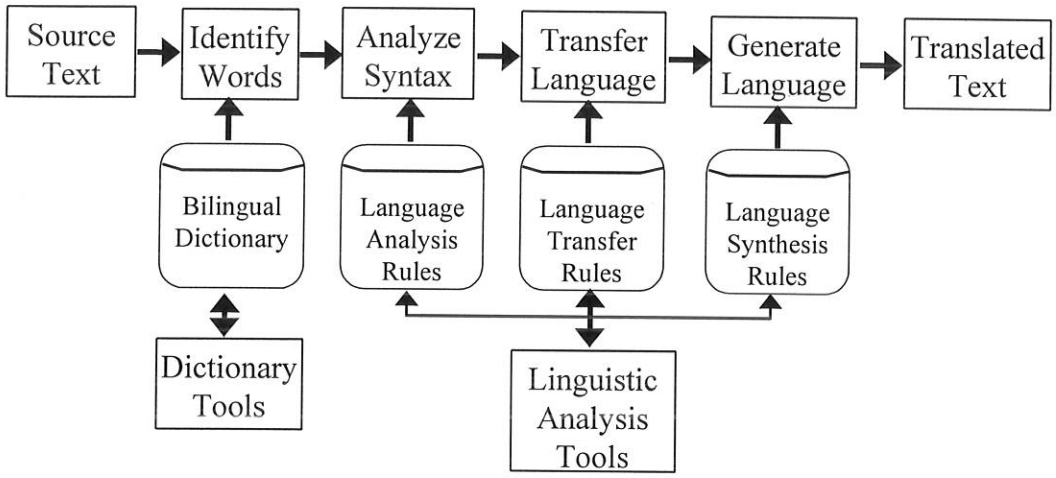

Figure 2.1: Schematic of document translation utilizing the transfer method

Source: Transparent Language Inc. (2000)

As can be seen in figure 2.1, the system accepts the source text and identifies the part of speech of each segmented word by referring the bilingual dictionary. The analysis stage then identifies the word form and the grammatical structure of each sentence aided by information about the meaning of the words in the sentence. Using the transfer rules, the elements of the sentence are reordered, inserted or deleted to form the syntax of the new language. Finally, the synthesis rule is the reverse step for analyzing the source text to generate the translated text according to the syntactic rules of the target language. (Transparent Language Inc. 2000)

2.2.2 Interlingual systems

The interlingual systems are also linguistics-based systems applying morphological and syntactic features but not semantic analysis of the input. However, some systems have features for derivation of semantic representations from the syntactic structure of expressions. Still others combine linguistic analysis and semantic-conceptual knowledge bases that involve understanding. (Hutchins 1994).

The Interlingual system would input the source language into an intermediate repository where expressions, sentence semantics and forms of expression would be replaced, independent of any language, and then the intermediate language, or meta-language, would be translated into the target language (Transparent Language Inc. 2000). Although, no convincing large-scale Interlingual MT system has yet been built, Maegaard (2000) approves that using the language-neutral analysis of text one can add new language at relatively low cost by creating only rules mapping from the new language into the Interlingual and back again.

In other words, as Smith (1990) explains, the goal of the analysis of the source text is to extract all the meaning from the input and is independent of the synthesis of the target text. There is a language independent intermediate form into which the source is translated and from which the output is generated.

2.3 Corpus-based approaches

The dominance of the rule-based systems in the 1980s has been broken by the emergence of the corpus-based systems since 1989. The principal feature of the corpus-based approaches is that no linguistic rules are applied. For example, the statistical and example-based approaches do not use syntactic or semantic rules in the analysis of texts or in the selection of lexical equivalent. (Hutchins 1994)

2.3.1 Statistical methods

Statistical methods were common in the 1960s but the results had been disappointing. Statistical methods are the sole means of analysis and generation without applying linguistic rules. Therefore, this method first aligns phrases, word groups and individual words of the parallel texts or translations of one another of any given two languages (Hutchins 1994). An alignment, according to Simard (1998), is a parallel segmentation of the two texts, into an equal number of segments, such that the n^{th} segment in one text and the n^{th} segment in the other text are translations of one another. The next task is to calculate the probabilities that any one word in a sentence of one language corresponds to a word or words in the translated sentence with which it is aligned in the other language (Hutchins 1994).

This method produces a result, for almost half of the phrases translated, either matched exactly the translations in the corpus, or expressed the same sense in slightly different words, or offered other equally legitimate translations. Although this method is meant to disprove the traditional rule-based approaches, it has been proposed to experiment with any methods especially the treatment of all

morphological variants of a single word and the use of syntactic transformations to bring source structures closer to those of the target language. (Hutchins 1994)

Besides, a pure statistics-based MT system according to Chen (1995) is usually incapable of processing long sentences and is usually domain dependent (that is, rules are not universal). Moreover, the computation time increases potentially with the length of sentences. However, such system is robust in processing partial and ill-formed sentences. Therefore, as Maegaard (2000) points out, the research on statistics-based MT systems did not convince the community that the statistics-only approach is the optimal path.

2.3.2 Example-based approaches

The 'example-based' also called the 'memory-based' approach was first proposed in 1984 but experiments began towards the end of the 1980s. The underlying hypothesis is that translation often involves the finding or recalling of analogous examples, the discovery or recollection of how a particular expression or some similar phrase has been translated before. The approach is founded on processes of extracting and selecting equivalent phrases or word groups from a databank of parallel bilingual texts, which have been aligned either by statistical methods or by more traditional 'rule-based' morphological and syntactic methods of analysis. (Hutchins 1994)

For calculating matches, some MT groups use semantic network or a hierarchy (thesaurus) of domain terms. Other groups use statistical information about lexical frequencies in the target language. The main advantage of the approach is that since the texts have been extracted from databanks of actual translations produced

by professional translators there is an assurance that the results will be accurate and acceptable. (Hutchins 1994)

However, Chen (1995) points out that the performance of an example-based system depends on the quality of collected examples and the similarity measure on examples and input sentences. When the matched units are phrase structures, the performance of such a system is better than that of a word-level system.

Finally, this approach, as Hutchins (1994) makes it clear, has been used most often to complement more traditional methods based on linguistic rules. Since example-based system is very effective in handling incomplete sentences or strange usage, a rule-based system is still required to handle complex grammatical sentences at the analysis and synthesis stages (Maegaard 2000).

2.4 Contents of machine-readable dictionaries

The contents of machine-readable dictionaries are highly dependent on the approaches followed and the complexity of the MT system described above. The Transparent Language Inc. (2000) describes the machine-readable dictionaries as the lexical backbone of all MT systems. They contain the bulk of the raw information by which text is translated. Each dictionary entry contains detailed grammatical and syntactical information about a source word and a target word. With this information, how words are used in a sentence are identified and replaced with grammatically and syntactically accurate foreign-language word equivalents.

The problem of such dictionaries is however the representations structure as well as the kind of information to be memorized. Humans can interpret dictionary contents in

a flexible and adaptive way, but the computer lacks such flexibility. Therefore, the dictionary must provide all detailed information needed for the MT system of particular interest. It may incorporate some or all of the following information.

2.4.1 Morphology

The morphological information enables the MT system to recognize a word in various forms. The word form of nouns for example, includes inflections, number, possession, gender, and so on. The word 'cows' is the plural form of the noun 'cow' and has a female gender. Therefore, as HEISOFT (2000) indicates, the morphology recognizes any form of a word or a phrase.

2.4.2 Syntax

Pang (1998) defines syntax as rules and constraints that are applied to words in order to form sentences. Information about the syntactic analysis determines the grammatical structure of each sentence. The syntax determines which word is the subject, the object, etc. Any grammatical errors, according to Pang (1998), are corrected by applying the syntax information to rearrange words.

Therefore, such syntax information, as Hutchins (1994) points out, are used for syntactic analysis of source texts to produce representations of both surface and deep (logical) relations of words in a given sentence. Surface-level information, according to Sakai (1998), is related to grammatical characteristics while deep-level information is associated with meanings.

2.4.3 Semantic

Semantic information determines the correct interpretation of a sentence from the multiple results produced by the syntactic analysis. Semantic information, as Hutchins (1994) describes, is limited to the identification of grammatical roles such as agent, instrument, etc. Maegaard (2000) further elaborates that the MT system must understand the actual meaning of the sentence. For instance, 'I am small' is expressed in many languages using the verb 'to be', but 'I am hungry' is often expressed using the verb 'to have', as in 'I have hunger'. Therefore, the dictionary must have information to incorporate such data about hunger.

Such semantic knowledge that is used to get unambiguous representation, according to Maegaard (2000), is therefore stored in a knowledge base dictionary. Without some level of semantic representation, MT systems will never be able to achieve high quality, because they will never be able to differentiate between cases that are syntactically ambiguous.

2.4.4 Concepts

He (1998), citing Larson (1984), defines concept as a recognizable unit of meaning in any given language. He (1998) further elaborates that in translation, a concept represented by a word in one language may be translated into a word, words, phrase, or even a sentence in another language. As an example can be the concept of 'a medical specialist who practices surgery' is represented by an English word 'surgeon' and by two Chinese words 'waike yisheng' representing for 'surgical

department' and 'doctor' respectively. Some concepts may be broader such like 'uncle' in English versus 'shushu' uncle from father's or mother's side in Chinese.

Further, a word may have varieties of meanings or concepts. Especially, for a word of a source language in a bilingual dictionary there exist many corresponding words and expressions in a target language. The variety of meanings between words of two languages may increase for languages that have different cultural, socioeconomic and technological backgrounds. Due to ambiguity of words or phrases, as HEISOFT (2000) explains, the MT system may not understand the input text and produces an incorrect meaning of the target language.

CHAPTER THREE

ENGLISH AND TIGRINYA NOUNS: A CONTRASTIVE ANALYSIS

3.1 Introduction

In Chapter two, it has been mentioned that machine-readable dictionaries contain detailed information about the source language as well as the target language. In order to get such information, the forms and functions of English and Tigrinya common nouns are discussed In this chapter. Common nouns are nouns that are not the name of a single particular person, place or thing. The similarities and differences of the forms and functions of nouns of both languages are discussed in detail. The functions of nouns include the gender, number, possessive, occurrence with articles, restrictions on word form (either to a singular or a plural form) and direct/indirect objects. The form of nouns indicates how nouns change their shapes due to affixations such as pluralization of a singular noun.

The information about English Nouns presented in this thesis is taken from the EDR home page (**http://iijnet.or.jp/edr/E_Struct.html**). The information regarding Tigrinya nouns is taken from Mason (1996), Palmer (1994), Tesfai (1993), Tsegay (1987), Girmay (1991), Girmay (1983), Bender et al (1976) and Leslau (1941).

3.2 Syllable

Before discussing the forms and functions of nouns of both languages, it is important to discuss the syllable. The information for syllable division of English nouns is useful to know whether the noun is hyphenated or not. Otherwise, a word divided by

a hyphen at a syllable division may not be found during dictionary lookup. The syllable division is also useful for determining how to break the English nouns at the end of a line of writing. The division between syllables is indicated by a slash (/). The syllable division marker for nouns that contain a slash as part of the notation is indicated by a double slash (//). No syllable division marking is inserted before or after a hyphen for words whose notation contains hyphens. Abbreviated words whose reading is the repetition of the alphabet letters are not given syllable division information. An example of the syllable division description used in English nouns is given below.

Noun	**Syllable division**
machine	ma/chine
x-ray	x-ray
FORTRAN	FOR/TRAN
CPU	CPU
king/queen	king//queen

In English, syllabic segment can be a vowel or consonant. For consonant syllabic segments, clusters of consonants are permitted word initially. A syllable can begin with a vowel or consonant and can be a single character or a word. For example, the underlined syllable of the noun 'con/fig/u/ra/tion' against 'farm'.

The syllable division and its application in Tigrinya language is however different. Unlike English, Tigrinya never uses hyphens to break nouns. The syllable

information is used only to show if the noun contains the correct vowel and consonant patterns according to the rules of Tigrinya language.

In Tigrinya, as Girmay (1983) points out, the syllabic segment is always a vowel and there is no syllabic consonant in this language. In other words, clusters of consonants are not permitted word initially. Girmay (1983) further elaborates that every syllable has to begin with a consonant but can end in a vowel or consonant. Thus, the syllable can be either of the two:

- CV - Single consonant followed by a single vowel (e.g. |do| 'Is it?')
- CVC - A single vowel between two consonants (e.g. |kab| 'from')

The syllable structure can be represented in a diagram as follows:

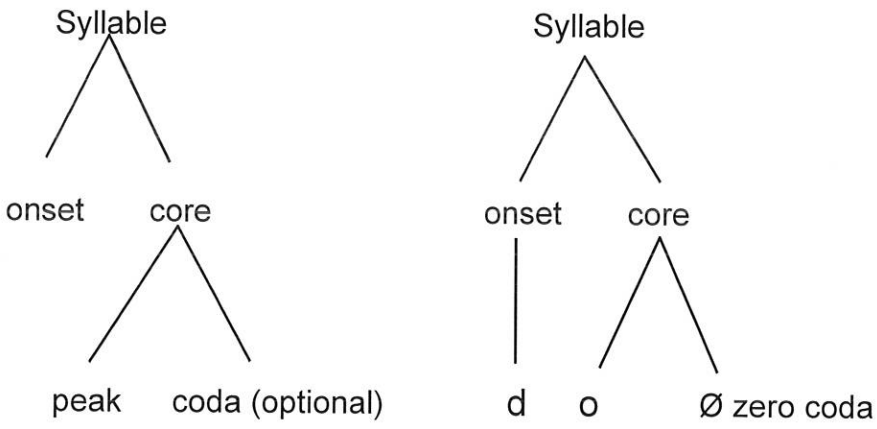

Figure 3.1: The structure of Tigrinya syllable

According to Girmay (1983), every syllable has a single consonant as onset, a single vowel as a peak and, optionally, a single consonant as coda. If clusters of two consonants occur word medial, the first member of the cluster goes with the first syllable, and the second member goes with the second syllable as:

|hafti| ⇨ |haf / ti| 'wealth'

CVC/CV

|haftam| ⇨ |haf / tam| 'rich'

CVC/CVC

Sometimes, dividing the consonant clusters to the nearest syllables may not be enough. It might need to insert the vowels 'i' and 'ɨ'. Both are epenthetic vowels. The vowel 'i' can be added to a word final, the vowel 'ɨ' can be added in word-medial position to break consonant clusters that are not permitted in Tigrinya language.

The epenthetic vowel 'ɨ' is inserted to break consonant clusters of loan words as:

|kläb| ⇨ |kɨläb| 'club'

or

The epenthetic vowel 'ɨ' can be inserted word medial and the epenthetic vowel 'i' can be added word final to noun forms that are composed of triconsonantal clusters that do not have any following vowels (Girmay 1991) as shown bellow.

|r?s| ⇨ | rɨ?si| 'head'

|brk| ⇨ |bɨrki| 'knee'

|?gr| ⇨ |?ɨgri| 'leg'

|s'fr| ⇨ |s'ɨfri| 'nail'

Without inserting the epenthetic vowel 'ɨ', initial consonant clusters appear. Without adding the epenthetic vowel 'i', final syllable can have a single consonant with out a vowel. In both cases, such consonant and vowel pattern with out inserting the epenthetic vowels is not permitted in Tigrinya syllable.

As Tigrinya does not permit initial consonant clusters, words beginning with the labio-velars seem they do not follow the syllabic rule. For example, the syllabic structure for 'g^{w}al' seems CCVC. But this is not the case. Girmay (1983) states that the labialised segments can be treated as having syllable structure CVCV (CuwV) and can be adopted as:

|guwal| ⇨ |g^{w}uwal| ⇨ |g^{w}ual| ⇨ |g^{w}al|

As has been discussed above, although viewing from different angles, the main point regarding syllable information of both languages is to help how nouns are correctly entered during the construction of the dictionary so that to get a correct analysis for input nouns and to correctly generate the various forms. The basis for noun formation and the noun form that bears the main meaning is discussed in the following section.

3.3 Noun formation

The basis for English word formation is a stem. Stem is the part of a word whose spelling more or less remains the same when different endings are added on to it. Stems are of different types. For example, man and wife are 'simple stems'. There are also 'compound stems' such as housemaid and snowfield. The last types of stems are 'complex stems' formed by the addition of derivational affix such as

womanliness and driver. All the mentioned stem types can add inflectional affixes to show grammatical relationships such as plural markers, possession, etc.

There are times when adding the inflectional suffix to stems can be done by removing or replacing some part of the stem. For the noun 'wife' for example, by adding the inflectional suffix 'ves' to the part of the stem 'wi', the plural noun 'wives' is formed. The part of the stem 'wi' can serve as an invariable portion of the stem to form both singular and plural. In English language therefore the stem carry of the basic meaning and the inflectional affixes carry the grammatical meaning.

The bases for Tigrinya word formation on the other hand are the consonantal skeletons called roots or radicals (Tesfai 1993, Bender et al 1976). The roots carry the basic lexical meaning and patterns of vowels carry the grammatical meaning. Words therefore are formed by changing the consonant and vowel patterns. For example, the verbal pattern /k'ätälä/ and the noun pattern /k'ätali/ are derived from the root /k'tl/. The stem for both verbal and noun patterns /k'ätäl/ constitute three tiers (Girmay 1991) as shown below:

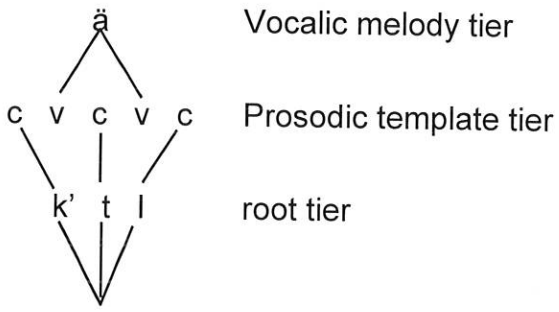

Figure 3.2: Three tiers for Tigrinya stem

The consonants as indicated by the root tier bear the basic meanings while the vowels form different patterns. Stems or simple words are built from consonantal

roots before other complex words are built from stems. That is, words are not only formed by derivation from roots, but also by affixation and compounding.

The bases for Tigrinya stems or words are therefore radical roots. For example, triradicals and quadriradicals have three and four consonants respectively at the root-tier level. Tigrinya radicals, according to Tesfai (1993) citing Anderson (1985) and McCarthy (1982), are three or four in number.

Nominal simple-stems formed from triradicals and the vowel patterns that can be used as the bases for derivational affixes are presented in Table 3.1.

Table 3.1: Simple nominal stems derived from triradicals

No	CV pattern	Examples	Vocalic pattern		UR of stem	SR of stem	Gemi-nation	Remarks
		Root	UR	SR				
1	ccc	frd	-	- ɨ - i	frd	fɨrdi		judgment
2	cccc	sbr	-	- ɨ - ɨ -	sbbr	sɨbbɨr	√	act of breaking
3	cvcvc	sʕl	- ä - a -	- ä - a - - a - a -	säʕal	säʕal saʕal		cough
4	cvcvc	ʕms'	- ä - ä -	- a - ä -	ʕämäs'	ʕamäs'		rebellion
5	cvcvc	ʕml	- ä - i -	- a - i -	ʕämil	ʕamil		customer
6	cvcvc	kdn	- i - a -	- i - a -	kidan	kidan		loyalty
7	cvccvc	ħfr	- ä - a -	- a - a -	ħäfar	ħafar	√	shy
8	cvccvc	rzn	- ä - i -	- ä - i -	räzzin	räzzin	√	heavy
9	ccvc	ml?	- u -	- ɨ - u -	mlu?	mɨlu?		full
10	ccvc	kdn	- a -	- ɨ - a -	kdan	kɨdan		clothe
11	cccvc	qds	- u -	- ɨ - u -	qddus	qɨddus	√	holy
12	cccvc	gmd	- a -	- ɨ - a -	gmmad	gɨmmad	√	piece
13	cvcc	ʕs'm	- ä -	- a - i -	ʕas'm	ʕas'mi		bone
14	cvcc	smʕ	- a -	-a - ɨ -	samʕ	samɨʕ		witness

Source: Tesfai (1993)

As indicated in table 3.1, all nouns have a triradical roots in common. However, they differ from each other in the vowel patterns, in gemination, in vowel and consonant arrangements, and so on. The Underlying Representation (UR) of stems is the basic form that matches with the Consonant Vowel (CV) pattern. However, if we look at such stems, we can see that the syllabic rule of the Tigrinya language is violated.

For example, the UR stems of numbers 1, 2 and 9 - 12 have an initial consonant cluster, which is not allowed in Tigrinya. For phonetic realization, the Surface Representation (SR) therefore inserts the epenthetic vowel |ɨ| in between the clusters. Likewise, the UR stems of the numbers 1, 2, 13 and 14 have consonant clusters word final. In the SR of stem, the epenthetic vowel |ɨ| is inserted in between the clusters of the word finals for the numbers 2 and 14 in order to make the final syllable CVC instead of CC; and since numbers 1 and 13 have a second syllable with only a consonant, the epenthetic vowel |i| is added at the end of the cluster, since |ɨ| is not allowed word final, to make the final syllable CV.

Furthermore, the vowel | ä | that immediately precedes or follows gutturals is changed to | a | as in numbers 3 - 5, 7 and 13. Quadriradicals can also be explained in the same way as triradical but they have a pattern difference by the number of consonants they contain. For example, the quadriradical stem |ħankas| 'lame' has a CV pattern of 'cvccvc', Root 'ħnks', but the same UR and SR of stem as the triradical |saçal| 'cough'.

Therefore, for English the stem and for Tigrinya the roots or radicals are the basis for word formation as well as for bearing the meanings of the words. The following sections discuss the functions and forms of both languages by taking the stems as the simple words to form complex nouns for both languages. The roots can be used when necessary for Tigrinya nouns.

3.4 Gender

The gender of English nouns can be masculine, feminine, neuter, or both masculine and feminine. Nouns that can be categorized as male are man, son, father, and so on. Nouns that are labeled as female are woman, daughter, mother and so on. There are also nouns that can be taken as either male or female such as student, child, etc. There are still other nouns that can be categorized as neutral such as book, teacher, etc.

Unlike English nouns, as Mason (1996) points out, most nouns in Tigrinya are either masculine or feminine. There is no neuter form. E.g. wädi 'boy', säbäyti 'woman', ?abo 'father', ?adä 'mother' or by custom s'ä*ħ*ay 'sun' (f.), wär*ħ*i 'moon' (f.), wäräk'ät 'paper' (m.), zär?i 'seed' (m.), n*ɨ*fas 'wind' (m.), maçs'o 'door' (m.), and so on. Some nouns, including most nouns describing inanimate objects can be used in either gender, e.g. gäza 'house', ?*ɨ*nsäsa 'animal' have their gender according to the context.

Besides, there are nouns in Tigrinya that add a suffix marker |-t| so that they become female nouns (Leslau 1941). Verbal nouns, according to Bender et al (1976), are marked for gender and most of the time the female marker in Tigrinya |-t| is added to a simple stem to form female nouns. The simple stems of verbal nouns are masculine and possess no special endings or have zero morphemes. For example:

Male	**Female**	
säbari	säbari-t	'one who breaks'
färadi	färadi-t	'judge'

ħaradi	ħaradi-t	'butcher'
s'äħafi	s'äħafi-t	'writer'
zäwari	zäwari-t	'driver'

There are also other nouns that change their gender in the same way as the above examples including for male noun marker |-wi| as well. For example, märça-wi for 'bridegroom' and märça-t for 'bride'.

Based on the above discussion, both languages have nouns that are masculine or feminine in gender by nature but verbal nouns in Tigrinya have gender marking for female and sometimes for male. Besides, there is no neuter form in Tigrinya. Grammatically, those nouns that are neutral in English are used as either male or female in Tigrinya.

3.5 Number (singular or plural forms)

Most English nouns add explicit suffixes to form their plurals. Some nouns add the plural suffix by replacing the final inflecting singular noun endings and others simply add the plural suffix to the singular nouns. Somehow they follow a regular inflection pattern. The plural suffixes are added to the invariable portion of the nouns, stems or to the singular noun forms. The invariable portion of the noun is not necessarily the stem of the word. For example, the invariable portion of the stem 'wife' which follows a regular inflection pattern (i.e. the 'fe' suffix ending becomes 'ves') is 'wi'. Nouns that follow a regular pattern to form their singular or plural are presented below:

Noun	Invariable portion	Singular noun-ending	Plural noun-ending
boy	boy	-	s
box	box	-	es
lady	lad	y	ies
wife	wi	fe	ves
leaf	lea	f	ves

Most English nouns can form their singular as well as plural forms as shown above. However, there are some nouns that do not follow any one of the above patterns and have irregular inflection patterns. For example, the plural form of the nouns 'man', 'tooth', is 'men', 'teeth' respectively.

There are also nouns that occur only in either singular or plural form. Abstract, concrete and abbreviated nouns like monopolism, water and USA respectively do not have a plural form. Nouns such as people, trousers, etc. have no singular form.

On the other hand if we look at the rules of plural formation in Tigrinya, according to Bender et al (1976), they are very complex. There are regular as well as irregular plural nouns. The regular plural nouns are formed by adding suffixes, prefixes or both; and/or by changing the internal structure of the nouns. Accordingly, Mason (1996) says that many Tigrinya nouns form their plural through the addition of a suffix.

One among the suffixes used to form a regular plural noun is the suffix |-at|. Nouns ending in 6th form are mostly changed to 4th form and add 't' (Mason 1996)[1]. Girmay (1983), Bender et al (1976) and Leslau (1941) also consider such nouns as having a consonant endings and add the suffix |-at|.

E.g.

t'ɨraz	'notebook'	t'ɨraz-at	'notebooks'
çarat	'bed'	çarat-at	'beds'
säb	'person'	säb-at	'persons'
näbiy	'prophet'	näbiy-at	'prophets'
çamät	'year'	çamät-at	'years'
?ɨsur	'prisoner'	?ɨsur-at	'prisoners'
s'ämam	'deaf'	s'ämam-at	'deaf people'

The other suffix to form plural nouns is the suffix |-tat|. As Mason (1996) describes, it is added to most Tigrinya nouns to form their plural. It is added, according to Girmay (1983), Bender et al (1976) and Leslau (1941), to nouns ending in vowels (ä, u, e and o). Most nouns ending with the vowel 'a' are also categorized in this group. The following are examples given by Leslau (1941):

[1] The vowel phonemes in Tigrinya are 'ä', 'u', 'i', 'a', 'e', 'ɨ', 'o' representing 1st, 2nd, 3rd, 4th, 5th, 6th, 7th forms respectively and 'ɨ' is optional.

E.g.

ħamu	'father-in-law'	ħamu-tat	'fathers-in-law'
däbdabe	'letter'	däbdabe-tat	'letters'
ʔadä	'mother'	ʔadä-tat	'mothers'
ʔabbo	'father'	ʔabbo-tat	'fathers'
ħaläx'a	'chief'	ħaläx'a-tat	'chiefs'
ʔɨmba	'mountain'	ʔɨmba-tat	'mountains'
gäza	'house'	gäza-tat	'houses'
ʂasa	'fish'	ʂasa-tat	'fishes'

However, nouns ending in 3rd form (i) need modification before adding the suffix |-tat|. Nouns ending in 3rd form must be changed to 6th form before adding |-tat|.

E.g.

kɨfli	'room'	kɨfli-tat	'rooms'
mɨdri	'land'	mɨdri-tat	'lands'
ʂaddi	'country'	ʂaddi-tat	'countries'

Nouns that end with 'a' (other than those nouns mentioned above that add the plural suffix |-tat|) and noun-endings such as 'ya' or 'ay' if the vowel 'a' immediately follows

or precedes the semi-vowel consonant 'y' are replaced by the suffix |-ot| to form their plurals (Leslau 1941).

E.g.

g^{w}asa	'shepherd'	g^{w}as-ot	'shepherds'
barya	'slave'	bar-ot	'slaves'
bɨs'ay	'companion'	bɨs'-ot	'companions'
ħarästay	'farmer'	ħaräst-ot	'farmers'

Although verbal nouns are discussed latter in this chapter, verbal nouns ending in 'i' are replaced only with the vowel |-o| such as:

E.g.

käfati	'one who opens'	käfat-o
säbari	'one who breaks'	säbar-o

The suffix female marker |-t| is added to form singular female verbal nouns to the above singular male verbal nouns. The plural form for male verbal nouns indicated above can also be used for plural female verbal nouns or it can simply be added the female marker to the plural male verbal noun. Most nouns derived from verbs therefore have more than one plurals as shown below (Tsegay 1987):

färadi	'judge'	färad-o, färad-ot, färäd-ti	'judges'
ħassawi	'liar'	ħassaw-o, ħassaw-ot, ħassäw-ti	'liars'

Among the three plural verbal noun forms indicated above, those indicated at the middle are the plural female verbal nouns and the rest are plural male or female verbal nouns.

Other than the verbal nouns, there are also some nouns that have two plural forms (Mason 1996).

E.g.

wäddi	'boy'	wädd-at	or ?a-wädd-at	'boys'
s~~i~~m	'name'	s~~i~~m-at	or ?a-s~~i~~m-at	'names'

In this case, both plural forms add the suffix |-at|. However, the second plural form also adds a prefix |?a-|. On the other hand, the following nouns unlike verbal nouns that end in 3rd form, form their plurals by adding the suffix |-tat| without dropping the noun ending or add the suffix |-at| by dropping the noun ending.

mäçalti	'day'	mäçalt-at	or mäçalti-tat	'days'
mägäddi	'path'	mägädd-at	or mägäddi-tat	'paths'
mäng~~i~~sti	'government'	mäng~~i~~st-at	or mäng~~i~~sti-tat	'governments'

As can be seen from the above examples, the plurals can be formed by suffixing the |-tat| or |-at|.

Although, they are not many in number, some nouns of Geez origin usually add the suffix |-an| (Leslau 1941).

E.g.

mämhɨr	'teacher'	mämhɨr-an	'teachers'
k'ɨddus	'saint'	k'ɨddus-an	'saints'
s'adɨk'	'just man'	s'adk'-an	'just men'
bɨs'uç	'the blessed'	bɨs'uç-an	'the blessed'

The other plural formation of Tigrinya nouns are the 'Internal' or 'broken' plurals formed essentially through changes inside the word rather than by suffixes. However, there is no resemblance in pattern between such plural nouns (Mason 1996). To some extent, such type of plural formation taken from Bender et al (1976) and Leslau (1941) has sub-classes such as:

Radical or Root	**Stem or singular**		**Plural**	
mnbr	mänbär	'chair'	mänabɨr	'chairs'
knfr	känfär	'lip'	känafɨr	'lips'
nbr	näbri	'tiger'	?a-nabɨr	'tigers'
nhb	nähbi	'bee'	?a-nahɨb	'bees'
gml	gämäl	'camel'	?a-gmal	'camels'
zmd	zämäd	'relative'	?a-zmad	'relatives'
?gr	?ɨgri	'leg'	?a-?gar	'legs'
?zn	?ɨzni	'ear'	?a-?zan	'ears'

E.g.

mämhɨr	'teacher'	mämhɨr-an	'teachers'
k'ɨddus	'saint'	k'ɨddus-an	'saints'
s'adɨk'	'just man'	s'adk'-an	'just men'
bɨs'uç	'the blessed'	bɨs'uç-an	'the blessed'

The other plural formation of Tigrinya nouns are the 'Internal' or 'broken' plurals formed essentially through changes inside the word rather than by suffixes. However, there is no resemblance in pattern between such plural nouns (Mason 1996). To some extent, such type of plural formation taken from Bender et al (1976) and Leslau (1941) has sub-classes such as:

Radical or Root	**Stem or singular**		**Plural**	
mnbr	mänbär	'chair'	mänabɨr	'chairs'
knfr	känfär	'lip'	känafɨr	'lips'
nbr	näbri	'tiger'	?a-nabɨr	'tigers'
nhb	nähbi	'bee'	?a-nahɨb	'bees'
gml	gämäl	'camel'	?a-gmal	'camels'
zmd	zämäd	'relative'	?a-zmad	'relatives'
?gr	?ɨgri	'leg'	?a-?gar	'legs'
?zn	?ɨzni	'ear'	?a-?zan	'ears'

ʔdg	ʔadgi	'donkey'	ʔa-ʔdug	'donkeys'
bʕry	bɨʕray	'bull'	ʔa-bʕur	'bulls'
k'mš	k'ämɨš	'dress'	k'ämawɨš	'dresses'
k'bʕ	k'obɨʕ	'cap'	k'äbawɨʕ	'caps'
drh	därho	'hen'	därahu	'hens'
ʕtr	ʕɨtro	'pot'	ʕataru	'pots'
ʕf	ʕif	'bird'	ʔa-ʕwaf	'birds'
ʔm	ʔom	'wood'	ʔa-ʔwam	'wood'
sm	sɨm	'name'	ʔa-sm-at	'names'
ħw	ħaw	'brother'	ʔa-ħw-at	'brothers'

As indicated above, broken plurals in Tigrinya can be formed by adding prefixes and/or suffixes in addition to the internal changes. Besides, the grouping above is based on the common CV pattern in forming plural nouns and the number of consonants at the root-tier level. However, the CV pattern and the type of vowels used to form the singular nouns with in the same group may differ in some cases. The plural formation in such cases should be from the root not from the stem. This is because there is vowel losing, vowel and/or consonant insertion and vowel change that occur during plural formation from singular forms.

With the exception of the nouns that have suffix forms |-tat| and |-at|, Girmay (1983) considers the rest nouns as non-productive and as applied to only a small percentage of the nouns in the language. Besides, the suffix |-at| is added to new words coming into the language and is productive more than the rest suffixes.

3.6 Possessive pronominals

Someone's possession of something is expressed by using the pronoun such as 'his', 'her', etc. to the thing possessed. In English, possessive pronominals are treated separately and are not attached to nouns. The possessive pronominals are my, our, your (s.), your (pl.), his, her, their. The possessive pronominals their, your (s.), and your (pl.) apply for both gender (male or female). All the possessive pronominals can be used with singular or plural nouns.

In the case of Tigrinya language, the possessive pronominals are attached to the nouns. Thus, according to Mason (1996), a noun can have many suffixes, a different one for each person, gender, and number. For example, the possessive forms for the noun gäza 'house' is presented as follows:

Person	Singular		Plural	
3m.	gäz-?u	'his house'	gäz-?om	'their house'
3f.	gäz-?a	'her house'	gäz-?än	'their house'
2m.	gäza-xa	'your house'	gäza-xum	'your house'
2f.	gäza-xi	'your house'	gäza-x~~i~~n	'your house'
1c.	gäza-y	'my house'	gäza-na	'our house'

The above examples indicate that the thing possessed 'house' is singular. The plural form for the noun gäza-wti 'houses' is presented as follows:

	Singular		Plural	
3m.	gäza-wt-u	'his houses'	gäza-wt-om	'their houses'
3f.	gäza-wt-a	'her houses'	gäza-wt-än	'their houses'
2m.	gäza-wt~~i~~-xa	'your houses'	gäza-wt~~i~~-xum	'your houses'
2f.	gäza-wt~~i~~-xi	'your houses'	gäza-wt~~i~~-x~~i~~n	'your houses'
1c.	gäza-wt-äy	'my houses'	gäza-wt~~i~~-na	'our houses'

Girmay (1983) describes the general rules to be followed at the boundaries of a stem to the possessed nominal. Before adding |-äy|, the possessive suffix for first person singular, the vowel 'ä' is deleted if the noun ends in vowels (o, ä, e and a) like biro - biro-y 'my office', bärbärä - bärbärä-y 'my pepper', k'~~i~~ddase - k'~~i~~ddase-y 'my prayer'. For nouns ending in vowel (u) however it is the 'u' that must be deleted as d~~i~~mmu -

dumm-äy 'my cat' and the vowel 'ɨ' is changed to 'u'. For nouns ending in (y or i), there must be a geminated consonant 'yy' before the possessive suffix as maħari - maħari-yyäy 'my saviour' and may - ma-yyäy 'my water'.

The rule for 3rd person masculine as well as feminine singular possessive is more or less similar. The possessive suffix for the former is |-u| and for the latter is |-a|. In both cases, the glottal stop '?' is inserted between the root ending and the suffix. For example, the changes as a result of the suffix |-u| are, dɨmmu - dɨmmu-?u 'his cat', dorho -dorhu-?u 'his hen', sɨrrä - sɨrru-?u 'his trousers', karra - karru-?u 'his knife', maħari - maħari-?u 'his saviour'. In this case, the noun endings (o, ä, a) are changed to 'u' as well. The changes that take place as the result of the suffix |-a| are, karra - karra-?a 'her knife', çilu - çilu-?a 'her donkey', maħari - maħari-?a 'her saviour', därho- därho-?a 'her hen', sɨrrä - sɨrrä-?a 'her trousers'

It is not different for feminine and masculine 3rd person plural possessives having the suffixes |-än| and |-om| respectively. Both insert the glottal stop '?' before the suffix. For example, karra - karra-?än and karra-?om 'their knife', därho - därho-?än and därho-?om 'their hen'.

Table 3.2 shows the summary of the possessive pronominal. A noun may end in one of the seven forms and the table indicates how such nouns are converted to form possessive pronominal using singular/plural person and female/male pronouns in combination with singular/plural-possessed things.

Table 3.2: Possessive pronominal endings

	Nouns ending in	my	our	your (m.s.)	your (f.s.)	your (m.pl.)	your (f.pl.)	his	her	their (m.pl.)	their (f.pl.)
5th	däbdabe	y	na	xa	xi	xum	xin	?u	?a	?om	?än
7th	maçs'o	y	na	xa	xi	xum	xin	?u	?a	?om	?än
2nd	ħamu	7th + y	na	xa	xi	xum	xin	?u	?a	?om	?än
4th	gäza	y	na	xa	xi	xum	xin	6th + ?u	6th + ?a	6th + ?om	6th + ?än
1st	?adä	y	5th + na	5th + xa	5th + xi	5th + xum	5th + xin	3rd + ?u	3rd +?a	3rd + ?om	3rd + ?än
3rd	çadi	1st + y	6th + na	6th + xa	6th + xi	6th + xum	6th + xin	2nd	4th	7th + m	1st + n
6th	mäs'ħaf	1st +y	na	xa	xi	xum	xin	2nd	4th	7th + m	1st + n

Source: Mason (1996)

3.7 Derivational nouns

Nouns derived from other part of speeches such as verbs, adjectives including from nouns themselves are called derivational nouns (Tesfai 1993, Tsegay 1987, Girmay 1983). The derivational nouns are treated separately because they add suffixes to stems of other parts of speeches to form nouns so that inflectional affixes such as plurals are added later. Inflectional affixation of the derivational nouns is done after the derivational suffix is added unlike the non-derivational nouns that add inflectional suffixes to singular noun stems. The following sub-topics discuss the derivational nouns.

In English there are suffixes that can be added to many verbs to form nouns. Nouns signifying the agent or the name of a profession are formed by adding the suffix |-er| to verbs and form nouns such as driver, fastener, opener, teacher, and so on. Abstract nouns on the other hand are formed by suffixing |-ment| to the verb and form nouns like amazement, development, payment, and retirement or |-ation| like admiration, association, examination and organization.

There are also derived nouns from adjectives. The suffixes added to adjectives to form nouns are |-ty|, |-ity| and |-ness|. Some of the examples are cruelty, oddity, and darkness.

In Tigrinya, many nouns are derived from verbs (Mason 1996). If we take the nouns s'ä*ħ*afi 'writer' and mäs'*ħ*af 'book', they are formed from a common stem s'ä*ħ*afä with a root meaning of 'write' which represents a verb (3rd person, singular and male).

There are two types of verbal nouns showing regular patterns as identified by Mason (1996) and Tsegay (1987): agentive and instrumentals.

The agentive nouns are nouns signifying the agent or the name of a profession usually formed with the vowels 1-4-3. Sometimes, they are formed with the vowels 1-4-4 plus 'y' as:

Verb		**Noun**	
säbärä	'he broke'	säbari, säbaray	'breaker'
färädä	'he judged'	färadi, färaday	'judge'
*ħ*assäwä	'he lied'	*ħ*assawi	'liar'
s'ä*ħ*afä	'he wrote'	s'ä*ħ*afi, s'ä*ħ*afay	'writer'

The second types of verbal nouns are nouns signifying tools or instrumentality formed with 'mə' plus 6-1-3 and sometimes plus 't' as mäs'räbi 'plane (carpenter's tool)', mäd*ħ*anit 'medicine' and mäfänk'äli 'sledg-hammer or wedge'. The following are some of the instrumental nouns:

Verb		**Noun**	
s'ärägä	'he cleaned'	mä-s'räg-i	'sth used for cleaning'
sätäyä	'he drank'	mä-stäy-i	'sth used for drinking'
sä?alä	'he drew'	mä-s?äl-i	'sth used for drawing'

As we can see, they all add prefix as well as suffix and there is also a vowel reduction. Similarly, the following have the same feature although they may differ in the suffix formation.

Verb		**Noun**	
gafäyä	'he cleaned'	mä-gaf-ya	'shovel'
zäwärä	''he walked'	mä-zawär-ya	'sth used by weavers'
däħanä	'he recovered'	mä-dħani-t	'medicine'
zäwärä	'he walked'	mä-zawäri-t	'vehicle'
säntärä	'he combed'	mä-säntär	'comb'
t'äħanä	'he ground'	mä-t'ħan	'mill'

Abstract nouns can also be derived from nouns. Nouns derived from substantives are abstract nouns expressing essence or quality as shown below:

Noun		**Derived form**	
ħaw	'brother'	ħiw-innät	'brotherhood'
mäħaza	'friend'	miħz-innät	'friendship'
leba	'thief'	leb-innät	'theft'
desa	'social'	des-innät	'socialism'
zämäd	'relative'	zimd-innä	'relationship'
g^{w}äräbet	'neighbour'	g^{w}irbit-innä	'neighborhood'

tɨzz	'reminiscence'	tɨzzɨ-ta	'remembrance'
harär	'eagerness'	harär-ta	'enthusiasm'
färäs	'horse'	färäs-äňňa	'horse-man'
bäräxa	'desert'	bäräx- äňňa	'desert-man'
läms'i	'leprosy'	läms'-am	'leper'
t'änk'i	'cause'	t'änk-am	'causative'

As we can see from the above, each group has its own distinguishing features such as suffixes, vowel pattern changes, vowel reduction and so on.

Nouns can also derive from adjectives. The following examples are derivation of nouns from adjectives:

Adjective		**Noun**	
ʕašša	'fool'	ʕɨšš-ɨnnät	'foolishness'
ħaffar	'shy'	ħɨffr-ät	'shyness'
sɨssuʕ	'greedy'	sɨssɨʕ-ɨnna	'greediness'
s'ɨnuʕ	'rigid/fixed'	s'ɨnʕ-at	'rigidity/research'

They have no common pattern but all lose the vowels of the original form.

As we can see in the above discussion, derived English and Tigrinya nouns are formed by adding suffixes to verbs, adjectives, and nouns. Both languages have similarities in

forming agentive nouns and derived nouns from adjectives. Besides, both apply more or less similar methods to form derived nouns by adding suffixes.

3.8 Direct/Indirect objects

In English, there is no special marker added to the noun. An object is said 'direct object' if it is directly affected by the action of the verb like the noun 'the money' in 'the man gave John the money'. An object is said 'indirect object' if it is an additional noun used after certain verbs which refers to the person or thing that an action is done to or for like the noun 'the students' in 'the teacher asked the students a question'.

Unlike in English, direct and indirect objects in Tigrinya are formed by prefixing 'nɨ' to the noun or object (Mason 1996, Palmer 1994, Bender et al 1976, Leslau 1941). When there are several objects, 'nɨ' can be prefixed to both objects or to the first object only.

näti ħaläx'an nätom wäthadäratɨn ħagäzom

or

nɨħaläx'an wäthadäratɨn ħagäzom

the chief / the soldier / he helped them

He helped the chief and the soldiers.

However, if the direct objects are addressed individually, it is better to use 'nɨ' with each of the nouns as 'nɨħaläx'an nɨwäthadäratɨn'.

The same is true to indirect objects as can be seen from the following example.

?ɨti č'ɨk'a nɨwäthadärat gänzäb hibwom

the/ chief/ to soldiers/ money/ he gave money to them

The chief gave money to the soldiers

3.9 Definite/indefinite articles

In English, the indefinite article has two manifestations: 'a' and 'an'. The singular countable nouns, countable groups and countable abstracts that begin with consonant sounds are preceded by the indefinite article 'a' and vowel sounds by the indefinite article 'an'. Countable nouns are nouns that can be separated and counted like books, students, and so on. Countable nouns have both singular and plural forms. Singular countable nouns must always have an article in front of it such as a, neither, this, each, and so on. Plural countable nouns can be used with both, some, more, many, etc. or with out a determiner. Plural countable nouns do not take indefinite articles.

Uncountable nouns on the other hand are nouns for things, which are thought as a mass or quantity like water, furniture, information, heat, and so on but not as separate singular or plural items. They can be used with determiners such as much, little, most, some, this etc but not with indefinite articles.

The other article is the definite article 'the'. Countable and uncountable nouns can be preceded by the definite article 'the'.

In Tigrinya like in English, the definite articles stand as a word and precede the nouns. But they differ in that there is no distinction between consonants as having vowel sound or consonant sound in Tigrinya. The other difference is that there is no indefinite article in Tigrinya. Moreover, different definite articles are used for male singular, male plural, female singular and female plural. The discussions below are presented based on Bender et al (1976) and Leslau (1941).

There are four definite articles. ʔɨtu or tu and ʔɨta are for singular while ʔɨtom and ʔɨtän are for plural male and female definite articles respectively.

Eg. ʔɨtu bɨɕray'the bull' ʔɨta laħmi 'the cow'

ʔɨtom ʔabɕur 'the bulls' ʔɨtän ʔalaħɨm 'the cows'

CHAPTER FOUR

DESIGN OF THE PROTOTYPE SYSTEM

4.1 Introduction

The previous chapter discusses the forms and functions of nouns in English and Tigrinya languages. In this chapter, the structure, the database design, the algorithms implemented for the experimentation and the result of the experimentation of the English-Tigrinya Bilingual Dictionary are discussed.

First, the structure of the database designed to store the English-Tigrinya Bilingual Dictionary is discussed. The structure of the database is based on the structure of the EDR and the analysis of English and Tigrinya nouns discussed in chapter three. Next, the design of the database based on the structure for storing the basic noun forms of both languages is discussed. Finally, the rules to form the inflected noun forms of both languages are implemented using algorithms, as a result, experimentation is conducted and the test results are discussed.

4.2 The structure of the English-Tigrinya Bilingual Dictionary

The structure and the database to be designed for storing a list of all complex forms of a word called lexicon depends on the nature of the language and the number of base forms to be included in the dictionary. Accordingly, both English and Tigrinya nouns must be taken into consideration.

If we look at the nature of both languages, they differ in how they add affixes as well as the number of the base forms. English has relatively simple inflectional morphology (Carter 1995). Thus, the set of affixes and their allowed combinations are manageable. Listing all inflected forms may be preferable.

Tigrinya however is a highly inflected language (Walther 1998, Tesfai 1993, Girmay 1983). Thus, according to Carter (1995), creating an exhaustive word list for highly inflecting languages is out of the question. The database therefore has to have a structure that incorporates the basic forms in addition to the list of roots/stems.

The structure of the English-Tigrinya bilingual dictionary is based on the structure of the Japanese Electronic Dictionary Research (EDR). There are two CD-ROMs called English-Japanese-Bilingual and English-Word Dictionaries that contain all information for English nouns. Unfortunately, the cost is too high to purchase even the single CD-ROM. Therefore, the codes for individual words are done manually with the help of the conventional paper-based dictionaries. The coding of the individual words is done by referring the coding system of the major forms and functions of English nouns incorporated in the documents available at the EDR home page. Codes proceeded by

the letter 'E' is for English nouns and by 'T' is for Tigrinya nouns. Although, the manual coding takes much time of the student researcher, a lot of effort has been put to include accurate English noun codes in the bilingual dictionary.

Accordingly, the English-Tigrinya Bilingual Dictionary is composed of the bilingual records arranged alphabetically according to the English headwords. The record of the English-Tigrinya Bilingual Dictionary includes: headword information, grammatical information, semantic information and management information. The structure of a record in the English-Tigrinya Bilingual Dictionary looks as shown in Table 4.1.

Table 4.1: Structure of English-Tigrinya Bilingual Dictionary Record

<Headword Information>	: (Section 4.2.1)
<Headword>	: Noun entries that resemble the entries in paper-based dictionaries
<Invariable portion of noun>	: Stem form of noun
<Root>	: For Tigrinya nouns only
<Adjacency attributes>	: Left and right adjacency
<Grammatical Information>	: (Section 4.2.2)
<word form restrictions>	: Occurrence of noun as singular or plural form only or its irregular form
<grammatical attributes>	: Countability, gender, collectivity, occurrence with articles
<Semantic Information>	: (Section 4.2.3)
<English head concept>	: English headword that represents

concept

<Tigrinya head concept> : Tigrinya headword that represents concept

<English concept explication> : Explanation of concept in English

<Tigrinya concept explication>: Explanation of concept in Tigrinya

<Management Information: : (Section 4.2.4) Date of entry or record modification

---

4.2.1 Headword information

The headword information contains alphabetically arranged English headwords. The headwords that resemble headword entries in conventional paper-based dictionaries facilitate the distinction between one word and another. Nouns that follow a regular inflection pattern such as the nouns 'dog' and 'car' are registered in their stem forms or in the singular forms. For some regular inflecting nouns, the portion of the noun that does not change during inflection like 'dictionar' of the noun 'dictionary' is also provided. Such invariable portion is used to form the final output forms of the nouns. All forms of nouns that do not follow regular inflection pattern like the noun 'man' and its plural form 'men' are registered as separate headword entries. Besides, nouns that follow a regular inflection pattern but whose meaning differs between the singular and the plural form like 'manner' and 'manners', are also registered as separate headword entries.

The headword information for Tigrinya nouns contains the Tigrinya headword. The Tigrinya headword is represented by headword that resembles headword entries in conventional paper-based dictionaries or the Surface Representation (SR) of stem. The root of the headword is also presented. When it is necessary, the root can be used as an input to generate the output (for broken plurals).

The headword information also includes adjacency attributes. They are given in the form of codes. They indicate the morphological restrictions to the adjacent words or constituents when the word is actually used. They indicate the possibility of joining morphemes to the left or to the right of the headword.

In English for example, the possessive pronominals, the definite and indefinite articles followed by blank space can be joined to the left of nouns. Similarly, in Tigrinya the definite articles followed by blank space and the object marker without blank space can be attached to the left of nouns.

Likewise, both English and Tigrinya nouns that follow regular inflection pattern can be followed by suffixes or right adjacency attributes or morphemes to form plurals. Nouns that do not follow regular inflection patterns have no right adjacency attributes. For instance, in English such nouns with final letter 's' such as analysis and not 's' such as man, tooth, datum are coded as a singular form without right adjacency since they do not involve regular suffix addition to form plurals and the plural forms with final letter 's' such as analyses, irises, matrices and not 's' men, corpora, mice are registered as separate headwords. In Tigrinya, there might also be plural forms not mentioned in this

thesis. If such noun that is not included in the discussion of Tigrinya plural forms is found, it is treated just like the English irregular nouns. Besides, the possessive pronominals in Tigrinya are part of the right adjacency attributes since they are suffixed to singular or plural nouns.

The left-adjacency and the right-adjacency attributes and codes used in the headword information are presented in Table 4.2 and Table 4.3 respectively.

Table 4.2: The left-adjacency attributes

Category	**Code**	**Example**
Noun with initial consonant sound	ELN1	man
Noun with initial vowel sound	ELN2	announcement
Singular male nouns	TLN1	bɨʕray
Singular female nouns	TLN2	laħmi
Singular male or female nouns	TLN3	gäza
Plural male nouns	TLN4	?abʕur
Plural female nouns	TLN5	?alaħɨm
Plural male or female nouns	TLN6	gäzawti

Note: No codes are provided for the English possessive pronominals and for Tigrinya object marker since they apply to all nouns without any restrictions with in the domain language.

Table 4.3: The right-adjacency attributes

Attribute	Code	Example	Comment/Explanation
Singular or plural forms	ECN1	book(s)	Noun followed by plural noun ending - s
	ECN2	dish(es)	Noun followed by plural noun ending - es
	ECN3	lad(y) lad(ies)	Noun followed by singular noun ending - y and plural noun ending - ies
	ECN4	kni(fe) kni(ves)	Noun followed by singular noun ending -fe and plural noun ending - ves
	ECN5	lea(f) lea(ves)	Noun followed by singular noun ending - f and plural noun ending - ves
	ERN1	man	Singular forms of nouns including irregular nouns
	ERN2	men	Plural forms of nouns including irregular nouns
	TCN1	säb(at)	Nouns followed by plural ending - at
	TCN2	ħamu(tat)	Nouns followed by plural

		ending - tat
TCN3	mɨdri	The 3rd form noun-ending is
	mɨdrɨ(tat)	changed to 6th form and add - tat
TCN4	mämhɨr(an)	Nouns followed by plural
		ending - an
TCN5	barya	The noun-ending ‘ya’ is
	bar(ot)	replaced by - ot
TCN6	bɨs’ay	The noun-ending ‘ay’ is
	bɨs’(ot)	replaced by - ot
TCN7	g^{w}asa	The noun-ending ‘a’ is
	g^{w}as(ot)	replaced by - ot
TCN8	käfati	The verbal noun-ending
	käfat(o)	‘i’ is replaced by - o
TCN9	nägaday	The verbal noun-ending
	nägad(o)	‘ay’ is replaced by - o
TCN10	mänbär	Quadriradical broken plurals
	mänabɨr	with VC pattern -ä-a- ɨ -
TCN11	näbri	Triradical broken plurals with
	?anabɨr	prefix ‘?a-’, VC pattern -a- ɨ -
TCN12	gämäl	Triradical broken plurals with
	?agmal	prefix ‘?a-’, VC pattern - -a-
TCN13	?adgi	Triradical broken plurals with
	?a?dug	prefix ‘?a-’, VC pattern - -u-

	TCN14	k'ämɨš	Triradical broken plurals with
		k'ämawɨš	VC pattern -ä-awɨ -
	TCN15	därho	Triradical broken plurals with
		därahu	VC pattern -ä(a)-a-u
	TCN16	?om	Biradical broken plurals with
		?a?wam	prefix '?a-', VC pattern -wa-
	TCN17	sɨm	Biradical broken plurals with
		?asmat	prefic '?a-', VC pattern - -,
			and suffix '-at'
	TRN1	gäza-wti	Singular forms of nouns including
			irregular nouns
	TRN2	hɨzbi	Plural forms of nouns including
			irregular nouns
Tigrinya	TPN1	?adä	Noun-ending in 1st form
possessives	TPN2	ħamu	Noun-ending in 2nd form
	TPN3	ʕadi	Noun-ending in 3rd form
	TPN4	gäza	Noun-ending in 4th form
	TPN5	däbdabe	Noun-ending in 5th form
	TPN6	mäs'ħaf	Noun-ending in 6th form
	TPN7	maʕs'o	Noun-ending in 7th form

Note: In the CV pattern, the consonant 'w' indicates its insertion and ä(a) indicates the vowel 'ä' can be changed to 'a' if it immediately follows or precedes a pharyngeal consonant.

4.2.2 Grammatical information

The grammatical information includes word form restrictions and grammatical attributes of both languages. Such information is given in the form of codes. The word form restriction indicates whether the headword occurs in the singular form only such as water, and monopolism or occurs in the plural form only such as corps. For irregular nouns mentioned in section 4.2.1, if they are singular nouns their plural headwords if they are plural nouns their singular headwords are entered here since the irregular form cannot be generated by codes. The grammatical attributes of the nouns include the countability, gender, collectivity and occurrence with articles. Collectivity indicates whether the noun being seen as a single collective entity, or as a collection of individual entities. Such nouns can be countable such as committee, army, etc or uncountable (having plural form only) such as people, trousers. Table 4.4 shows the grammatical information.

Table 4.4: The grammatical information

Attribute	**Code**	**Comment/Explanation**
Word form restrictions	ENRQSG	Noun occurs in the singular form only
	ENRQPL	Noun occurs in the plural form only
	man/men	The singular form for plural headwords or vice versa for irregular nouns
	TNRQSG	Noun occurs in the singular form only
	TNRQPL	Noun occurs in the plural form only
	NULL	Nouns that follow regular inflection pattern
Countability	ENC	Countable

	ENU	Uncountable
	TNC	Countable
	TNU	Uncountable
Collectivity	ENG	Collective noun (Ex. people)
	TNG	Collective noun (Ex. ḥizbi)
Gender	ENM	Referent of noun is male (Ex. man)
	ENF	Referent of noun is female (Ex. woman)
	ENNE	Referent of noun is neutral (Ex. book)
	ENCM	Referent of noun is either male or female (Ex. Student, baby)
	ENCA	Referent of noun is either male, female or neutral (Ex. baby)
	TNM	Referent of noun is male (Ex. säb?ay)
	TNF	Referent of noun is female (Ex. säbäyti)
	TNCM	Referent of noun is either male or female (Ex. tämhari)
Co-occurence with articles	ENNODEF	Nouns that do not occur with definite article
	ENNOINF	Nouns that do not occur with an indefinite article
	ENNOAR	Nouns that do not occur with an article
	TNNODS	Nouns that do not occur with singular definite article
	TNNODP	Nouns that do not occur with plural definite article
	TNNOAR	Nouns that do not occur with definite article

4.2.3 Semantic information

Under the semantic information, English and Tigrinya head concepts and concept explications are included. A concept is the image or idea that comes to one's mind by looking or hearing a word. Concepts provide the information necessary to discriminate between the various meanings of the nouns contained in the English-Tigrinya Bilingual Dictionary. Head concepts are described in both English and Tigrinya by a headword that best represents the concept. In addition, a written explanation of the concept called concept explication is provided in order to help people distinguish the various concepts of a noun. For example, for the headword 'dog' there can be at least two head concepts and two concept explications. The first head concept is 'dog' in English and 'xälbi' in Tigrinya and the concept explication is 'an animal called dog'. The second head concept for the headword 'dog' in English is 'spy' and in Tigrinya 'sälayi' to represent the concept explication 'a person who spy'. In such cases, there can be more than one headword registered in the dictionary for a single noun having different head concepts and concept explications.

4.2.4 Management information

The management information contains information related to date of record creation or date of record update.

4.3 Design of the English-Tigrinya Bilingual Dictionary

Based on the structure just discussed, the database that contains the basic noun forms of both languages is designed having the structure as shown in table 4.5.

Table 4.5: Attributes and data types of the English-Tigrinya Bilingual Dictionary

Attribute	Type	Size	Remark
English - headword	Character	20	
Tigrinya - headword	Character	20	
English - invariable portion	Character	20	
Tigrinya - root	Character	10	
English - left adjacency	Character	5	
Tigrinya - left adjacency	Character	5	
English - singular and plural form	Character	5	
Tigrinya - plural form	Character	5	
Tigrinya - possessive	Character	4	
English - word form restriction	Character	20	Can be Null
Tigrinya - word form restriction	Character	5	Can be Null
English - countability	Character	3	
Tigrinya - countability	Character	3	
English - collectivity	Character	3	Can be Null
Tigrinya - collectivity	Character	3	Can be Null
English - Gender	Character	4	
Tigrinya - Gender	Character	4	
English - co-occurrence with articles	Character	7	Can be Null
Tigrinya - co-occurrence with articles	Character	7	Can be Null
English - Head concept	Character	20	
Tigrinya - Head concept	Character	20	
English - concept explication	Character	50	
Tigrinya - concept explication	Character	50	
Date updated	Character	10	

4.4 Algorithms for translating nouns and the experiment

The previous section presents how the database has been designed and what type of data to be included. Based on the database design, a sample list of English and Tigrinya nouns and coded noun forms are entered (see Appendix 1). Next, there have to be rules to form all inflected forms of both languages. This section therefore first discusses the language synthesis rules or algorithms necessary to form the different noun forms of both languages. The overall process needed to convert the varying forms of a source noun to a target noun is shown in the following diagram.

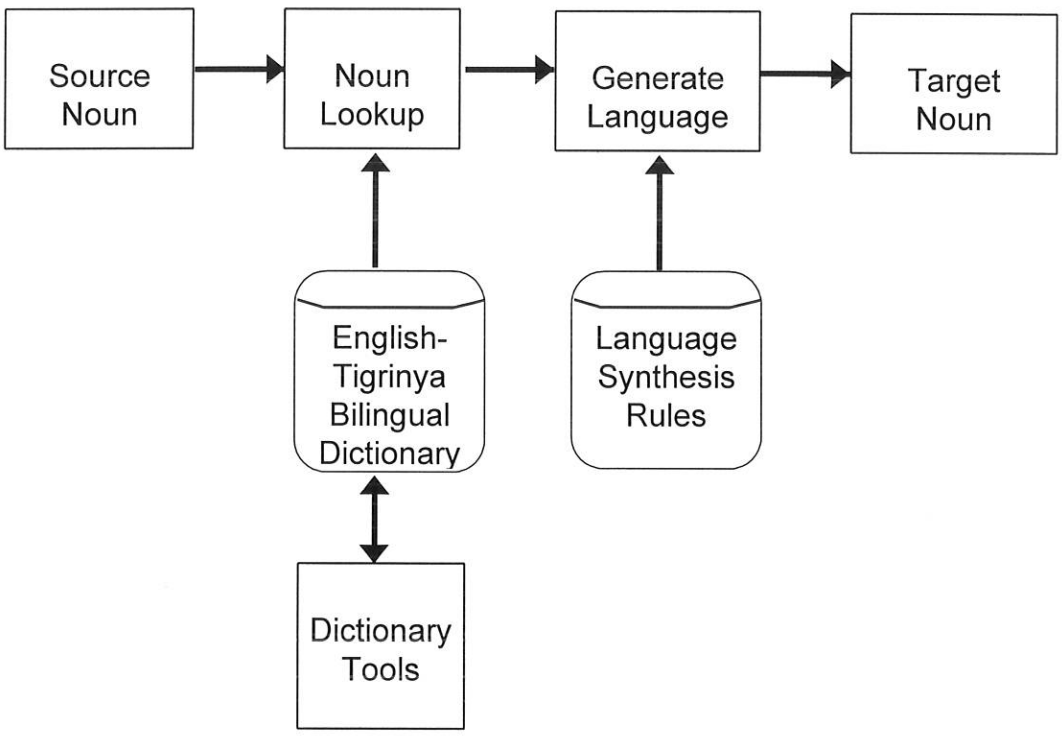

Figure 4.1: An overview diagram for translating English and Tigrinya nouns

In the first process, English or Tigrinya noun stem is accepted as an input by the system. Note that the stem forms of the nouns are the same as the noun entries in conventional paper-based dictionaries. Besides, for generation purpose, for English the invariable portion of the noun and for Tigrinya either the stem or the root is the basis. As the stem noun is accepted, it is checked or looked up in the English-Tigrinya Bilingual Dictionary database. The information containing the various forms for the source noun and its equivalent target noun is already incorporated in the database with the help of dictionary tools such as English and Tigrinya paper-based dictionaries. Then, the source or input noun and the target noun are synthesized side by side using the language synthesis rules of both languages before the final target noun form is generated. The generation or synthesis rules are therefore discussed in the following sub-topics.

4.4.1 Affixes to nouns

In the discussion of machine translation in chapter two, the synthesis rule is a rule for generating the different word forms. The various noun forms of both languages as discussed in chapter three are formed by the addition of prefixes and suffixes. Anderson (1992) states that to address the question of placing the inflectional formatives of a word in their correct relation to one another, the set of rules are subject to a language-particular ordering. That is, the fact that a word has the form stem + affix1 + affix2 reflects the rule attaching affix1 applied before the rule that attaches affix2 of a particular language. It is important to discuss the usage of noun affixes to form the different noun forms of both languages.

The English nouns, as discussed in chapter three, have only singular and plural endings (suffixes) that are added to the invariable portion of the regular nouns. To form plurals for regular nouns, the singular noun-endings are replaced by the plural noun-endings or vice versa to form singulars. Table 4.6 shows the singular and plural noun-endings used to form singular or plural noun forms in English.

Table 4.6: Noun-endings for regular English nouns

No	S1	S2	Example
1	-	s	boy boys
2	-	es	box boxes
3	y	ies	lad(y) lad(ies)
4	fe	ves	wi(fe) wi(ves)
5	f	ves	lea(f) lea(ves)

Where:

S1: Singular noun ending

S2: Plural noun ending

Other than the noun-endings discussed above, the definite and indefinite articles can precede English nouns.

Tigrinya nouns on the other hand have affixes for number, gender, possessive and object as indicated in Table 4.7. Since an affix is not replaced by another like the English nouns discussed above, the order of the affixes is important in order to form varying Tigrinya noun forms.

Table 4.7: Affix orders to Tigrinya nouns

No	P1	P2	S1	S2	S3
1	nɨ	?a	t	at	äy
2				tat	na
3				ot(o)	xa
4				an	xi
5				wti	xum
6					xɨn
7					u
8					a
9					om
10					än

Where:

P1: Object marker

P2: Prefix for plural

S1: Female marker

S2: Suffix for plural

S3: Possessive marker

The combination of the prefixes and the suffixes are presented in Table 4.8 and Table 4.9 respectively.

Table 4.8: The combination of noun prefixes in Tigrinya

No	Prefix	Example
1	P1	nɨ-?om
2	P1P2	nɨ-?a-?wam
3	P2	?a-?wam

Table 4.9: the combination of noun suffixes in Tigrinya

No	Suffix	Example
1	S1	s'ä*ħ*afi-t
2	S1S3	s'ä*ħ*afi-t-äy
3	S2	färad-o
4	S2S1	färad-o-t
5	S2S3	mämhɨr-an-äy
6	S3	mämhɨr-äy

In addition to the prefixes and suffixes mentioned above, Tigrinya nouns are also preceded by definite articles.

Based on the above discussion, the algorithm developed for the different affixes are discussed next.

4.4.2 Developing algorithms for affixes

As indicated in the above discussions, the rules for applying the affixes are important in generating the varying noun forms of both languages. As indicated in diagram 4.1, the generation of the varying noun forms is done based on the data stored for affixes in the bilingual dictionary. Using these data, the generations of the various noun forms of both languages are processed. Accordingly, the algorithms for processing number, gender, possessive and article are discussed below.

For English and Tigrinya nouns, data used for plural noun formation is stored in the bilingual dictionary. The algorithm for processing plural noun forms is presented below.

Figure 4.2: The algorithm for processing English and Tigrinya plural noun forms

---

Stem_noun = invariable portion of a noun
Noun = noun that resemble paper-based dictionary headword entry
Suffix = plural ending
Plural = the plural noun form
Database = the English-Tigrinya Bilingual Dictionary

0. Start
1. Read Noun
2. If Noun is plural
    Return Plural = Noun
3. If Noun is singular
    3.1 If Noun occurs in singular form only, goto 4
    3.2 If Noun is irregular noun
        Return Plural = the plural form stored in the database, goto 4

3.3 Return Plural = Stem_noun + Suffix

4. End

The algorithm first accepts a noun that resembles a noun entry in paper-based dictionaries such as nouns entered in singular or plural, regular or irregular, etc. form. If the noun is plural, further process is not performed. If the noun is singular however it is tested further whether it occurs in singular form only or whether it has a plural form. Nouns that occur only in singular form, no plural form is generated. If the noun has a plural form and follows a regular inflection pattern to form plural noun, then for English nouns, the noun ending is added to the invariable portion of the noun stored in the database. For Tigrinya nouns, the plural suffix is added to the stem or to the singular form or to the singular female verbal noun. For nouns that do not follow regular inflection pattern, it is difficult to generate the plural form and the plural form stored in the database is generated.

The next algorithms are used to process the definite and indefinite articles. The algorithms used to process the articles for English and Tigrinya nouns are not the same. They have their own features as discussed in chapter three and as indicated in the structure of the bilingual dictionary.

The following algorithm is used to process the indefinite article for English nouns.

Figure 4.3: The algorithm for processing indefinite article for English nouns

---

Singular = Singular noun forms

Plural-SF = Plural nouns that have also singular form

Noun = noun that resemble paper-based dictionary headword entry

Sing = Singular form of Noun

0. Start

1. Read Noun

2. If Noun is Singular or Plural-SF

2.1 If Noun occurs with indefinite article

2.1.1 If Noun begins with consonant sound

Return Indefinite = "a " + Sing

2.1.1 if noun begins with vowel sound

Return Indefinite = "an " + Sing

3. Else Return Indefinite = ""

4. End

First, a noun is accepted and checked if it is singular or plural. If the noun is singular or if it is plural that has singular form, then it is checked whether it occurs with indefinite article or not. If the noun occurs with indefinite article, the noun is preceded by the indefinite article "a" or "an" followed by a blank space for noun that begins with consonant sound or vowel sound respectively. Singular nouns that do not take indefinite article and plural nouns that do not have a singular noun form do not have a noun form preceded by indefinite article. This algorithm works only for English nouns since Tigrinya nouns do not have indefinite articles.

The other article is the definite article. The algorithms for processing definite article for English and Tigrinya nouns are discussed below. The following algorithm is for processing singular and plural definite articles for English nouns.

Figure 4.4: The algorithm for processing definite article for English nouns

---

Singular = Singular noun forms
Plural= Plural noun forms
Noun = noun that resemble paper-based dictionary headword entry

0. Start
1. Read Noun
2. If Singular
    2.1 If occurs with definite article
        Return Definite = "the " + Singular
3. If Plural
    3.1 If occurs with definite article
        Return Definite = "the " + Plural
4. Else Return Definite = ""
5. End

The algorithm for the definite article can be applied for singular or plural and for regular and irregular noun forms. If the accepted regular or irregular noun is singular noun and can be preceded by the definite article, the definite article is inserted before the singular noun. The same is true for plural nouns. The definite article for plurals can be processed after the plural noun formation algorithm is processed as indicated in figure

4.2. Singular or plural nouns that do not occur with definite article are not preceded by the article.

The following algorithm is for processing Tigrinya nouns in relation to definite articles.

Figure 4.5: The algorithm for processing definite article for Tigrinya nouns

---

Singular = Singular noun forms

Plural= Plural noun forms

Noun = noun that resemble paper-based dictionary headword entry

0. Start

1. Read Noun

2. If Singular

2.1 If occurs with definite article

2.1.1 If Singular is female

Return Definite = "ʔɨta " + Singular

2.1.2 If Singular is male

Return Definite = "ʔɨtu " + Singular

3. If Plural

3.1 If occurs with definite article

3.1.1 If Plural is female

Return Definite = "ʔɨtom " + Plural

3.1.2 If Plural is male

Return Definite = "ʔɨtän " + Plural

4. Else Return Definite = ""

5. End

The algorithm for processing the definite article for Tigrinya nouns shows that the definite article preceded to the noun depends on the gender in addition to the number of the noun. Singular female, singular male, plural female and plural male nouns are

preceded by different definite articles. Thus, Tigrinya nouns are checked for number, gender and if they occur with definite article or not. Nouns that can be treated as female or male can be preceded by definite articles representing both genders.

The next algorithm works for Tigrinya possessives. The algorithm for English possessive pronominals is not included here since they can be simply preceded to any singular or plural nouns.

Figure 4.6: An algorithm for Tigrinya possessive pronominals

Noun = singular or plural noun
Suffix = possessive pronominal endings
0. Start
1. Read noun
2. For Suffix = 'y' / my
2.1 If noun has an ending vowel 'ä' or 'a' or 'e' or 'o'
Return Possessive = noun + Suffix
2.2 Else if noun has an ending vowel 'u'
Return Possessive = noun with changed vowel ending to 'o' + Suffix
2.3 Else if noun has an ending vowel 'i'
Return Possessive = noun with deleted vowel ending + 'ä' + Suffix
2.4 Else
Return Possessive = noun + 'ä' + Suffix
3. For Suffix = 'na' or 'xa' or 'xi' or 'xum' or 'xin' /our, your (m.s.), your (f.s.), your (m.pl.), your (f.pl.) respectively
3.1 If noun has an ending vowel 'u' or 'a' or 'e' or 'o' or consonant

Return Possessive = noun + Suffix

3.2 Else if noun has an ending vowel 'ä'

Return Possessive = noun with changed vowel ending to 'e' + Suffix

3.3 Else

Return Possessive = noun with changed vowel ending to 'ɨ' + Suffix

4. For Suffix = 'u' or 'a' or 'om' or 'än' /his, her, their (m.pl.), their (f.pl.) respectively

4.1 If noun has an ending vowel 'u' or 'e' or 'o'

Return Possessive = noun + '?' + Suffix

4.2 Else if noun has an ending vowel 'ä'

Return Possessive = noun with changed 'ä' to 'i' + '?' + Suffix

4.3 Else if noun has an ending vowel 'a'

Return Possessive = noun with changed 'a' to 'ɨ' + '?' + Suffix

4.4 Else if noun has an ending vowel 'i' or consonant

Return Possessive = noun with changed ending vowel to 'u' for the possessive 'u', and to 'a' for the possessive 'a', to 'o' + 'm' for the possessive 'om', to 'ä' + 'n' for the possessive 'än'

5. End

The algorithms discussed above are coded using Microsoft Visual C++ and the dictionary is constructed using Microsoft Access. The next section presents the test results.

4.4.3 The experiment and the test result

The test is conducted to see whether the data stored in codes in the English-Tigrinya Bilingual Dictionary is correctly coded and can be used for translation purposes. Although, the major aim of the thesis is to build the stated machine-readable dictionary, the test being conducted helps to see whether the dictionary can be used as an input for machine translation from English to Tigrinya and vice versa.

The test is conducted using a prototype English-Tigrinya Bilingual Dictionary that contains data regarding articles, gender, collective and number for nouns of both languages. The headwords for English and Tigrinya nouns are selected that represent:

- Nouns that follow regular inflection pattern to form plural.
- Nouns that follow regular inflection pattern but whose meaning differs between the singular and plural form.
- Nouns that do not follow regular inflection pattern to form plural.
- Countable and uncountable nouns.
- Plural noun forms for irregular nouns.
- Collective nouns that can be treated as singular or plural.
- Nouns having multiple concepts.
- Singular nouns having no plural form.

- Plural nouns having no singular form.
- Nouns that occur with definite and/or indefinite articles.
- Nouns that do not occur with definite and/or indefinite articles.
- Nouns with different genders.
- Nouns with initial consonant or vowel sound for English nouns.

The test is conducted to see the outcomes of the above listed concepts and the similarities or differences that might occur in translating from English to Tigrinya or vice versa using the data stored in the database and the algorithms developed.

Accordingly, 50 headwords for English nouns and their Equivalent Tigrinya headwords and 50 headwords for Tigrinya nouns and their equivalent English headwords, all in all 200 sample noun headwords are manually entered to the English-Tigrinya Bilingual Dictionary Database using Microsoft Access (see Appendix 1).

Using the sample nouns and the algorithm discussed for processing the indefinite article, definite article and plural, the following result is achieved (see Appendix 2).

- The equivalent noun forms for both languages are displayed.
- Nouns that can be preceded by the definite and indefinite articles are displayed.
- Singular and plural nouns that do not occur with definite or indefinite articles are correctly expressed by the terms 'has no indefinite', 'has no singular definite' and 'has no plural definite'.
- The plural forms for regular and irregular nouns are displayed.
- Singular nouns that do not have plural forms and vice versa are treated correctly.
- Some times, collective nouns in English can be either singular or plural in Tigrinya or vice versa. For example, the noun 'choir' in English is a collective noun having both singular and plural form. Its equivalent Tigrinya noun is 'mäzämiran', which is plural. Such variation is possible to incorporate in the bilingual dictionary.
- The equivalent for plural nouns can also be singular or vice versa. For example, the equivalent Tigrinya noun for the English noun 'people' is 'hiżbi'. The former is plural while the latter is singular. Such types of noun forms are also incorporated in the dictionary.
- Nouns that follow regular inflection pattern but whose meaning differs between the singular and plural form are correctly displayed. For example, the noun 'fathers' with the concept 'any of the people from whom somebody is descended' neither occurs with articles nor has singular form and it differs from the noun 'father' with

the concept 'any male animal in relation to its young' having a singular and plural form that can occur with articles.

- During data entry, some headword nouns in English are encountered as having no equivalent headword nouns in Tigrinya. A related noun that more or less can represent the English noun is given with some modifications in the concept explication. For example, for the noun 'boy' there are two noun headings in Tigrinya 'wäddi' or 'män?ɨsäy'. Either of them are not the exact translation for the stated English noun. The former noun stands for the English noun 'male' and the latter for the noun 'youngster'. Therefore, the noun 'wäddi' is taken by modifying the concept explication to 'male youngster'.

- For some English headword nouns, the equivalent Tigrinya headword nouns can be adjectives. In such cases, the wrong result is produced. For example, the English noun 'prisoner' is '?ɨsur' in Tigrinya. It can be treated as having both genders in Tigrinya. However, the singular female form becomes '?ɨsurti', which does not follow the rules in female noun formation in Tigrinya. Instead of adding the female marker 't' it adds 'ti'. In such cases, an alternative noun form is entered to solve the problem. Thus, the other alternative Tigrinya headword 'ḥɨbus' is taken.

- For Tigrinya singular headword nouns having broken plurals have been given their own codes for each sub categories as is given to regular headword nouns. The intended input to form the plural for such nouns was the radical. However, it is not only difficult to construct the plural form from the radical but also the listed broken

plurals in chapter three are not by any means exhaustive. To simplify such complications, all broken plurals including the ones not discussed in this thesis are treated as irregular nouns. Thus, the Tigrinya broken plurals can be treated in the same way as the English irregular nouns with out further addition of attribute and with out developing further complicated algorithms.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

The purpose of the thesis is to construct an English-Tigrinya Bilingual Dictionary using the tools and techniques used for developing the EDR dictionary. In order to build the bilingual dictionary, information regarding noun forms and functions of both languages has been discussed. The dictionary is developed based on the noun forms and functions of the two languages. In order to test how the constructed database can be applied for machine translation purpose, selected sample headwords of both languages are entered to the database. Translation of English noun forms to Tigrinya noun forms is performed using these data and the algorithms developed to process the varying noun forms of both languages. The algorithm is developed for testing whether the coded information can be used for translation purpose and to find the possible problems. In this chapter, the summary of the thesis and recommendations for further research is presented.

5.2 Conclusion

Although, Tigrinya nouns have complex inflecting noun forms that are not present in English, the EDR system can be used for constructing the English-Tigrinya Bilingual Dictionary with minor changes. The minor changes can be reflected in the approaches followed to store the Tigrinya possessive pronominals and broken plurals.

The basic forms for Tigrinya possessive pronominals, for example, can be represented by a single attribute. If such possessive pronominals are attached to nouns, there can be more than thirty noun forms: ten for male, ten for female and more than ten for plural nouns having two or three plurals. The English-Tigrinya Bilingual Dictionary by using the EDR system can incorporate such type of varying noun forms with a single attribute. Thus, it becomes easier to manage a bilingual dictionary that has a limited number of attributes.

It is also possible to reduce the complexity of Tigrinya noun form by using the techniques of the EDR. The complexity of broken plurals in Tigrinya, for example, is resolved by applying the techniques applied for Irregular plural forms in English.

Besides, Tigrinya has prefixes for object noun forms, which are not available in English. All direct or indirect nouns are prefixed with out any limitations and can be incorporated in the algorithms rather than in the bilingual dictionary.

The process of generating noun forms is language dependent. For both languages, language dependent rules for suffix and prefix orders as well as rules for suffix and

prefix addition to basic noun forms are applied. The algorithms that are used to generate the different noun forms therefore can apply the rules.

Both languages have similarities in some cases. The similarity can be seen in the way the definite article is preceded to the singular or plural noun forms in both languages. The other similarity is that plural forms are generated by adding suffixes to basic noun forms of both languages. In such cases, a single algorithm with some modification can be applied for both languages.

Finally, using the techniques and tools used to build the English-Tigrinya Bilingual Dictionary for noun forms, there is also a possibility to build the other parts of speech such as verbs and adjectives.

5.3 Recommendations

The prototype English-Tigrinya Bilingual Dictionary accepts nouns similar to noun entries used in paper-based dictionaries. Nouns other than these are not recognized such as the plural form of regular nouns unless the nouns are registered in the dictionary as plural forms for irregular nouns or the nouns have only plural forms. In order to recognize and accept such nouns, a stemmer or a parser is useful if developed to work with the dictionary.

Adding data to the English-Tigrinya Bilingual Dictionary using the manual tools is the most difficult task. Especially, data for gender, number, article etc. for Tigrinya nouns is almost unavailable in the currently available paper-based dictionaries. Further research has to be conducted for applying techniques so that the bilingual dictionary can learn the different noun forms and provide options for newly added headword entries.

Rules that treat the vowel changes due to insertion, deletion or change for Tigrinya nouns, especially to broken plurals and possessive pronominals have to be further studied.

The EDR is developed for example based MT systems. However, in order to test an example-based MT system, the bilingual dictionary must incorporate as well the other parts of speech such as verbs, adjectives, preposition, etc. This is because, a fully-fledge MT system only works at least at sentence level. Therefore, those interested can develop bilingual dictionaries for the mentioned part of speech in Tigrinya.

The constructed dictionary can also be applied to other local languages.

Finally, other features not included in this study can also be identified and studied by others who have interest in such study.

Bibliography

Academy of Ethiopian Languages. (1975 E.C.). "Yä Tɨgriñña mäzgäbä k'alat zɨgjɨt_yä t'ɨnat wäräk'ät". Yä mäzgäbä k'alat simpoziyäm t'äqlala riporɨt. 173-199 Addis Ababa: Academy of Ethiopian Languages.

_____. (1989 E.C.). Mäzgäbä k'alat Tɨgriñña-bɨ Tɨgriñña. Addis Ababa: Academy of Ethiopian Languages.

Adult Education Department. (1975 E.C.). "Yä tɨmhirɨt mäsariyawočɨn bäyäbɨheräsäb k'^{w}ank'^{w}awoč yä mätärgom lɨmɨd". Yä tɨrgum seminar Gɨnbot 14-16 1972 E.C. 331-346. Addis Ababa: Bahɨlna Sɨporɨt Guday Comišɨn.

Anderson, S.R. (1992). A-morphous morphology. Cambridge: Cambridge University Press.

Bender, M.L. et al. (Eds) (1976). Language in Ethiopia. London: Oxford University Press.

Carter, D. (1995). "Rapid development of morphological descriptions for full language processing systems". Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics, March 1995. 202:209. Dublin: Association for Computational Linguistics.

Chen, Kuang-hua and Hsin-Hsi Chen. (1995). Machine translation: an integrated approach. Taipei: National Taiwan University.

Coulbeaux, P.S. et J. Schreiber. (1915). Dictionnaire de la langue Tigraï. Wien: Alfred Holder.

Curriculum Department. (1972 E.C). "Yä mastämariya mäs'haftin bämazägajät bäkul yätädärägu t'irätočina yagat'ämu čigroč". Yä tirgum seminar Ginbot 14-16 1972 E.C. 300-307. Addis Ababa: Bahilna Siporit Guday Comišin.

Electronic Dictionary Research (EDR). **http://iijnet.or.jp/edr/E_Struct.html**.

Eritrean People's Liberation Front (EPLF). (1986). Tigrinya-English dictionary. Rome: Research and Information Center on Eritrea.

Girmay Berhane. (1983). The phonology of Tigrigna. Thesis for the degree of master of arts in linguistics. Addis Ababa: Addis Ababa University.

_____. (1991). Issues on the phonology and morphology of Tigrinya. Doctoral dissertation. Montreal: University du Quebec.

He, S. (1998). "Concept similarity and conceptual information alteration via English-to-Chinese and Chinese-to-English translation of medical article titles". Journal of the American Society for Information Science. 49(2): 169-175.

HEISOFT. (2000). What is language software?

http://www.cs.cmu.edu/~ref/mlim/chapter4.html

http://www.cs.cmu.edu/~ref/mlim/chapter4.html

http://www.heisoft/sprachse.htm.

Hutchins, J. (1994). Research methods and system designs in machine translation: a ten-year review, 1984-1994. Norwich: University of East Anglia.

Kahaner, D.K. (1992). Report: Electronic Dictionary Research Institute (EDR). URL: **http://www.iijnet.or.jp/edr/J_index.html**.

Kinfe Regeb Zeleqe. (1972 E.C.). "Yä ?astärg^{w}aminät muya bä?ityopiya hibrätäsäb tarik". Yä tirgum seminar Ginbot 14-16 1972 E.C. 134-181. Addis Ababa: Bahilna Siporit Guday Comišin.

Leslau, W. (1941). Documents Tigrigna. Paris: Librairie C. Klincksieck.

Maegaard, B. Ed. (2000). Machine translation.

Mason, J. (1996). Tigrinya grammar. New Jersey: The Red Sea Press.

Mesfin Ghebrehiwet. (1996). A self-study Tigrinya for foreigners and English for Tigrinya speakers. S.l.: Sn.

Oriolo, L., Amanuel Sahle, Senait Iyob. (1997). How to say it in English, Tigrinya, Italian. Asmara: the Red Sea Press.

Palmer, F.R. (1994). Grammatical roles and relations. New York: Cambridge University Press.

Pang, S. (1998). Introduction to machine translation: an online tutorial. **http://geocities.yahoo.com/toto?s=76000008**

Research and Information Center of Eritrea. (1986). Tigrinya-English dictionary. Rome: Research and Information Center of Eritrea.

Sakai, S. (1998). Introduction to IT projects (Electronic Dictionary Research: EDR). Tokyo: Japan Electronic Dictionary Research Institute.

Simard, M. (1998). The BAF: a corpus of English-French bitext. Montreal: Université de Montréal.

Smith, P.D. (1990). An introduction to text processing. Cambridge: the MIT Press.

Solomon Deressa. (1969). "Translation - a necessary art? Addis Reporter weekly magazine. 1(23): 18-21.

Tesfai Tewolde. (1993). Word formation in Tigrinya. Addis Ababa: Addis Ababa University.

Transparent Language Inc. (2000). Machine translation and the power and purpose of 'gisting' in the Internet era: a white paper from transparent language. **http://www.transparentlanguage.com/ets/about/mtwhitepaper.htm**

Tsegay Tafere. (1987). Derivation of nouns in Tigrigna. For the Degree Bachelor of Arts. Addis Ababa: Addis Ababa University.

Tuquabo, A. (1987). Concise English-Tigrinya dictionary. Asmara: Ethiopian Studies Center.

Uqbamicael Habtemariam. (1993). English-Tigrinya dictionary. 1st Ed. S.l.: HAGOS.

Walther, M. (1998). Declarative prosodic morphology online - Tigrinya verbs. **http://pc0864.germanistik-kunst.uni-marburg.de/~walther**

Wedekind, C. (1972 E.C). "Which translation in which language". Yä tirgum seminar Ginbot 14-16 1972 E.C. 308-330. Addis Ababa: Bahilna Siporit Guday Comišin.

Yohanes G/egziabher. (1949). Mäzgäbä qalat Tigriñña-?amharñña. Asmara: Artigrafit Eritrea.

Appendix 1: Database record structure and sample data

Headword	Stem	LA	RA	Restrict	Coun	Col	Gender	CoWA	HeadCon	ConExp	DateUpd
analysis	analysis	eln2	ern1	analyses	enc		enne		analysis	the study of something by examining its parts and relationships	11/06/2001
tìntane	tntn	tln3	tcn2		tnc		tncm		tìntane	gìls'i nay mìgbar kunätat	
army	arm	eln2	ern1	army	enc	eng	encm		army	a military force concerned with attack and defence on land	08/06/2001
särawit	srwt	tln1	trn1	tnrqsg	tnc	tng	tnm		särawit	bìzìHi zäläwo nìx'inat zìtädaläwä ?as'war zìHazä säb	
audience	audience	eln2	ern1	audience	enc	eng	encm		audience	the people listening or watching a performance, speech	08/06/2001
sämäAi	smA	tln1	trn1	tnrqsg	tnc	tng	tnm		sämäAi	?admas'i hìzbi	
band	band	eln1	ern1	band	enc	eng	encm		band	a group of people formed for some common purpose	08/06/2001
ganta	gnt	tln2	tcn2		tnc		tnf		ganta	Hadä Aìlama zäläwom säbat zìtä?akäbìlu gujlä	
basket	basket	eln1	ecn1		enc		enne		basket	a container for holding or carrying things	10/06/2001
zämbil	zmbl	tln3	tcn1		tnc		tncm		zämbil	bìfìHso laxa zìsìraH zìnt'ìlt'äl mätHaz ?ax'ìHa	
bed	bed	eln1	ecn1		enc		enne		bed	a piece of furniture for sleeping on	08/06/2001
Aarat	Art	tln3	tcn1		tnc		tncm		Aarat	säb zìdìk'ìsälu nawti gäza	
belief	belief	eln1	ecn1		enc		enne	ennoar	belief	a religion or something taught as part of a religion	10/06/2001
?ìmnät	?mnt	tln3	trn1	tnrqsg	tnu		tncm	tnnoar	?ìmnät	?ab Hadä haymanot mätkäl Hasab mìxìtal	
book	book	eln1	ecn1		enc		enne		book	a number of printed sheets fastened together within a cover	10/06/2001
mäs'Haf	s'Hf	tln3	trn1	mäs'aHìfti	tnc		tncm		mäs'Haf	fidäl zìtäx'äräs'älom wäräx'ax'ìti ?ab Hadä zìtät'äräzu	
box	box	eln1	ecn2		enc		enne		box	a container made of wood or metal with a flat base and sides	11/06/2001
sas'un	ss'n	tln1	tcn1		tnc		tnm		sandux'	?arbaAìt ?ìgri mä?azìn zäläwo bìAìns'äyti wäy Has'in zìsìraH	
boy	boy	eln1	ecn1		enc		enm		boy	a young male person	11/06/2001
wäddi	wd	tln1	trn1	wäddat	tnc		tnm		wäddi	täbaAtay zìx'onä wìlad säb	
calf	cal	eln1	ecn5		enc		encm		calf	the young of a cow	12/06/2001
mìrax	mrx	tln3	tcn1		tnc		tncm		mìrax	wädi laHmi	
cattle	cattle	eln1	ern2	enrqpl	enc	eng	encm	ennoinf	cattle	cows and bulls kept on farms for milks and meat	10/06/2001
käbti	kbt	tln5	trn2	tnrqpl	tnu	tng	tncm		käbti	?alahìm ?abAur bìHabära zìs'ìwìAulu sìm	
choir	choir	eln1	ern1	choir	enc	eng	encm		choir	group of people who sing together during religious services	08/06/2001
mäzämìra	zmrn	tln6	trn2	tnrqpl	tnc	tng	tncm		mäzämìran	?ab betä kìrìstìyan bìzema zägälgìlu säbat	

Headword	Stem	LA	RA	Restrict	Coun	Col	Gender	CoWA	HeadCon	ConExp	DateUpd
chorus	chorus	eln1	ern1	chorus	enc	eng	encm		chorus	group of people who sing together	08/06/2001
mäzämìra	zmrn	tln6	trn2	tnrqpl	tnc	tng	tncm		mäzämìran	hìdì? zìbälä zema zìs'awätu säbat	
clan	clan	eln1	ern1	clan	enc	eng	encm		clan	group of families originally descended from the same family	08/06/2001
Aalet	alt	tln1	trn1	tnrqsg	tnc	tng	tnm		Aalet	nay sìga zìmìdìna zäläwom Hadä Aaynät hìzbi	
clergy	clergy	eln1	ern2	enrqpl	enc	eng	encm	ennoinf	clergy	the members of the priesthood that perform religious services	10/06/2001
kahnat	khn	tln4	trn2	tnrqpl	tnu	tng	tnm		kahnat	k'äši	
clothes	clothes	eln1	ern2	enrqpl	enu	eng	enne	ennoinf	clothes	garments such as trousers that are worn to cover the body	12/06/2001
kìdan	kdn	tln1	trn1	kìdawnti	tnc	tng	tnm		kìdan	säbat zìgäbruwo zìtäsäfäyä Aaläba	
earnings	earnings	eln2	ern2	enrqpl	enu	eng	enne	ennoinf	earnings	money which is earned by working	12/06/2001
?ìtot	?tt	tln1	trn1	tnrqsg	tnu		tnm		?itot	nay dìxam nay sìraH waga	
father	father	eln1	ecn1		enc		enm		father	any male animal in relation to its young	09/06/2001
?abbo	?b	tln1	tcn3		tnc		tnm		?abbo	täbaAtay wäladi	
fathers	father	eln1	ern2	enrqpl	enu	eng	enne	ennoar	ancestor	any of the people from whom somebody is descended	09/06/2001
wälädi	wld	tln4	trn2	tnrqpl	tnu	tng	tnm		tìwlìddi	Haräg wäläd o	
fish	fish	eln1	ecn2		enc		enne		fish	a creature that lives in water	09/06/2001
Aasa	As	tln3	tcn2		tnc		tncm		Aasa	?ab baHri zìnbìr	
foot	foot	eln1	ern1	feet	enc		enne		foot	movable part of the body on which a person or animal stands	12/06/2001
?ìgri	?gr	tln3	trn1	?a?gar	tnc		tncm		?igri	säb konä ?insìsa zìxädälun t'ät'äw zìblälun kìfili ?akal	
fox	fox	eln1	ecn2		enc		encm		fox	a small dog-like flesh-eating wild animal	10/06/2001
wäxarya	wxry	tln3	trn1	wäxaru	tnc		tncm		wäxarya	kìndi kälbi ?ìtìxäwìn t'eläbägiA ?ìtìbälìA ?arawit	
girl	girl	eln1	ecn1		enc		enf		girl	a young woman who is not married	10/06/2001
gwal	gwl	tln2	trn1	?agwalat	tnc		tnf		gwal	zäytämärAawät dìngìl	
knife	kni	eln1	ecn4		enc		enne		knife	a blade fixed in a handle used for cutting as a tool or weapon	12/06/2001
karra	kr	tln3	trn1	kararu	tnc		tncm		kara	nìmäHìrädi mäxtäfi zägälgìl bìHas'in zìsìraH ?ax'Ha	
lady	lad	eln1	ecn3		enc		enf		lady	a woman with dignity and good manners	11/06/2001
wäyzäro	wyzr	tln2	trn1	wäyzärazìr	tnc		tnf		wäyzäro	nìzìtämärAwät säbäyti zìwähab mäAarìg	
leaf	lea	eln1	ecn5		enc		enne		leaf	green flat parts of a plant growing from a stem or branch	11/06/2001
k'os'li	k's'l	tln3	tcn3		tnc		tncm		k'os'li	?ab ?a?wam zìbäk'ul k'ät'älìya Hìbri zäläwo ?akal	

Headword	Stem	LA	RA	Restrict	Coun	Col	Gender	CoWA	HeadCon	ConExp	DateUpd
life	li	eln1	ecn4		enc		enne	ennoinf	life	the active force in animals and plants	12/06/2001
hìywät	hywt	tln3	trn1	tnrqsg	tnu		tncm	tnnoar	hìywät	säb ?ìnsäsa täxli wäzätä mìnbar zìx'ì?ìlulu Hayli	
loaf	loa	eln1	ecn5		enc		enne		loaf	a mass of bread shaped and baked in one piece	12/06/2001
Hìnbaša	Hnbš	tln3	trn1	Hanabìš	tnc		tncm		Hìnbaša	bìHìrus' sìrnay zìsìnkät rìgud zìbälä mägbi	
louse	louse	eln1	ern1	lice	enc		enne		louse	small wingless insect that live on the skin and hair	12/06/2001
x'umal	xml	tln1	tcn1		tnc		tnm		x'umal	däm säb ?ìndamäs'äyä zìnäbìr balìA	
man	man	eln1	ern1	enrqsg	enu	eng	enne	ennoar	human race	the human race	10/06/2001
säb	sb	tln1	trn1	tnrqsg	tnu	tng	tnm	tnnoar	wädi säb	zär?i säb zäbälä fit'ur kulu	
man	man	eln1	ern1	men	enc		enm		man	an adult male human being	10/06/2001
säb?ay	sb?y	tln1	trn1	säb?ut	tnc		tnm		säb?ay	bäs'Hi wädi täbaAtay	
manner	manner	eln1	ern1	enrqsg	enu		enne		manner	a personal way of acting or behaving towards other people	12/06/2001
t'äbay	t'by	tln1	trn1	tnrqsg	tnu		tnm		t'äbay	s'ìbux' yìxun Hìmax' bahrìy säb zìs'ìwäAalu kal	
men	man	eln1	ern2	man	enc		enm		man	an adult male human being	10/06/2001
säb?ut	sb?y	tln4	trn2	säb?ay	tnc		tnm		säb?ay	bäs'Hi wädi täbaAtay	
mother	mother	eln1	ecn1		enc		enf		mother	any female animal in relation to its young	09/06/2001
?ìnno	?n	tln2	tcn3		tnc		tnf		?ìnno	wäladit	
news	news	eln1	ern1	enrqsg	enu		enne	ennoinf	news	fresh information or reports of recent events	10/06/2001
zena	zn	tln3	trn1	tnrqsg	tnu		tncm		zena	?ìwanawi wärä	
objective	objective	eln2	ecn1		enc		enne		objective	an aim	10/06/2001
Aìlama	Alm	tln3	tcn2		tnc		tncm		Aìlama	kìtìbäs'ìHo ?ìtìs'ìArälu nägär	
ox	ox	eln2	ern1	oxen	enc		enm		ox	a fully-grown male of the cattle family	12/06/2001
bìAray	bAry	tln1	trn1	?abAur	tnc		tnm		bìAray	nìmaHräs zìAaxälä täbaAtay käbti	
patch	patch	eln1	ecn2		enc		enne		patch	small piece of material used to cover a hole or damaged place	10/06/2001
mälgìba	lgb	tln3	tcn2		tnc		tncm		mälgìba	zìtäx'ädädä kìdan ?ìnìlägbälu	
people	people	eln1	ern2	enrqpl	enc	eng	enne		people	persons in general	12/06/2001
hìzbi	hzb	tln1	tcn2		tnc	tng	tnm		hìzbi	bìzHi zäläwom säbat	
prisoner	prisoner	eln1	ecn1		enc		encm		prisoner	a person or animal held with limited freedom of movement	09/06/2001
?ìsur	?sr	tln3	tcn1		tnc		tncm		Hìbbus	zìtätHazä säb	

Headword	Stem	LA	RA	Restrict	Coun	Col	Gender	CoWA	HeadCon	ConExp	DateUpd
shelf	shel	eln1	ecn5		enc		enne		shelf	a flat long and narrow board for putting or storing things on	12/06/2001
käbHi	drdr	tln1	tcn2		tnc		tnm		käbHi	?ax'Hu mäs'aHìfti ?ìnäx'ämìt'älu	
star	star	eln1	ecn1		enc		enne		star	distant bodies appearing as a point of light in the sky at night	10/06/2001
koxob	kxb	tln3	trn1	käwaxìbti	tnc		tncm		koxob	nay gäza? rì?su bìrhan zäläwo ?ab t'äfär zìrkäb	
team	team	eln1	ern1	team	enc	eng	encm		team	group of people who work or play together	08/06/2001
ganta	gnt	tln2	tcn2		tnc		tncm		ganta	?ab zìxonä s'äwäta bìHadä xoynom mìskal?ot zìgät'mu säbat	
thief	thie	eln1	ecn5		enc		encm		thief	a person who steals	12/06/2001
leba	lb	tln3	tcn2		tnc		tncm		leba	zäynatu zìwäsìd	
tooth	tooth	eln1	ern1	teeth	enc		enne		tooth	small hard bony objects growing inside the mouth	12/06/2001
sìnni	sn	tln3	trn1	?asnan	tnc		tncm		sìnni	mìs gìrs'an zìtätäHaHazä nìmägbi mäHyäxi zìtäk'ìm	
troop	troop	eln1	ecn1		enc		encm		troop	group of soldiers	08/06/2001
särawit	srwt	tln1	trn1	tnrqsg	tnc	tng	tnm		särawit	bìzìHi zäläwo nìx'inat zìtädaläwä ?as'war zìHazä säb	
water	water	eln1	ern1	enrqsg	enu		enne	ennoinf	water	a liquid that falls as rain, is in lakes, rivers, seas	12/06/2001
may	my	tln1	trn1	tnrqsg	tnu		tnm		may	?ab Aila baHri k'älay zìrkäb Hìbrìn mäk'ärätìn zäybìlu zìsìtä	
wife	wi	eln1	ecn4		enc		enf		wife	the woman to whom a man is married	10/06/2001
säbäyti	sbyt	tln2	trn1	?anìšti	tnc		tnf		säbäyti	bäAälti säb?ay	
wolf	wol	eln1	ecn5		enc		encm		wolf	a wild animal of the dog family which hants other animals	12/06/2001
täxula	txl	tln3	tcn3		tnc		tncm		täxula	zär?i kälbi zìxonä ?ab dur zìnäbìr tägada?i ?arawit	
youth	youth	eln1	ecn1		enc		enm		youth	a young man	08/06/2001
män?ìsäy	n?s	tln1	tcn1		tnc		tnm		män?ìsäy	Aìdmi?u bìzuH zäykonä säb	

Appendix 2: Translated sample noun forms of both languages

Headword	ConExp	Singular	Indefinite	S-Definite	P-Definite	Plural
analysis	the study of something by examining its parts and relationsh	analysis	an analysis	the analysis	the analyses	analyses
tìntane	gìls'i nay mìgbar kunätat	tìntane	tìntane	?itu tìntane, ?ita tìntane	?itom tìntanetat, ?itän tìntanetat	tìntanetat
army	a military force concerned with attack and defence on land	army	an army	the army	the army	army
särawit	bìzìHi zäläwo nìx'inat zìtädaläwä ?as'war zìHazä säb	särawit	särawit	?itu särawit	Has no plural definite	Has no plural
audience	the people listening or watching a performance, speech	audience	an audience	the audience	the audience	audience
sämäAi	?admas'i hìzbi	sämäAi	sämäAi	?itu sämäAi	Has no plural definite	Has no plural
band	a group of people formed for some common purpose	band	a band	the band	the band	band
ganta	Hadä Aìlama zäläwom säbat zìtä?akäbìlu gujlä	ganta	ganta	?ita ganta	?itän gantatat	gantatat
basket	a container for holding or carrying things	basket	a basket	the basket	the baskets	baskets
zämbil	bìfìHso laxa zìsìraH zìnt'ìlt'äl mätHaz ?ax'ìHa	zämbil	zämbil	?itu zämbil, ?ita zämbil	?itom zämbilat, ?itän zämbilat	zämbilat
bed	a piece of furniture for sleeping on	bed	a bed	the bed	the beds	beds
Aarat	säb zìdìk'ìsälu nawti gäza	Aarat	Aarat	?itu Aarat, ?ita Aarat	?itom Aaratat, ?itän Aaratat	Aaratat
belief	a religion or something taught as part of a religion	belief	Has no indefinite	the belief	Has no Plural Definite	beliefs
?ìmnät	?ab Hadä haymanot mätkäl Hasab mìxìtal	?ìmnät	?ìmnät	Has no singular Definite	Has no plural definite	Has no plural
book	a number of printed sheets fastened together within a cover	book	a book	the book	the books	books
mäs'Haf	fidäl zìtäx'äräs'älom wäräx'ax'iti ?ab Hadä zìtät'äräzu	mäs'Haf	mäs'Haf	?itu mäs'Haf, ?ita mäs'Haf	?itom mäs'aHìfti, ?itän mäs'aHìfti	mäs'aHìfti
box	a container made of wood or metal with a flat base and side	box	a box	the box	the boxes	boxes
sas'un	?arbaAìt ?ìgri mä?azìn zäläwo bìAìns'äyti wäy Has'in zìsìra	sas'un	sas'un	?itu sas'un	?ìtom sas'unat	sas'unat
boy	a young male person	boy	a boy	the boy	the boys	boys
wäddi	täbaAtay zìx'onä wìlad säb	wäddi	wäddi	?itu wäddi	?ìtom wäddat	wäddat
calf	the young of a cow	calf	a calf	the calf	the calves	calves
mìrax	wädi laHmi	mìrax	mìrax	?itu mìrax, ?ita mìrax	?ìtom mìraxat, ?ìtän mìraxat	mìraxat
cattle	cows and bulls kept on farms for milks and meat	Has no sing	Has no indefinite	Has no singular definite	the cattle	cattle
käbti	?alahìm ?abAur bìHabära zìs'ìwìAulu sìm	Has no sing	Has no singular	Has no singular Definite	?itän käbti	käbti
choir	group of people who sing together during religious services	choir	a choir	the choir	the choir	choir
mäzämìran	?ab betä kìrìstìyan bìzema zägälgìlu säbat	Has no sing	Has no singular	Has no singular Definite	?itom mäzämìran, ?itän mäzämìra	mäzämìran

Headword	ConExp	Singular	Indefinite	S-Definite	P-Definite	Plural
chorus	group of people who sing together	chorus	a chorus	the chorus	the chorus	chorus
mäzämìran	hìdì? zìbälä zema zìs'awätu säbat	Has no sing	Has no singular	Has no singular Definite	?ìtom mäzämìran, ?ìtän mäzämìra	mäzämìran
clan	group of families originally descended from the same family	clan	a clan	the clan	the clan	clan
Aalet	nay sìga zìmìdìna zäläwom Hadä Aaynät hìzbi	Aalet	Aalet	?ìtu Aalet	Has no plural definite	Has no plural
clergy	the members of the priesthood that perform religious servic	Has no sing	Has no indefinite	Has no singular definite	the clergy	clergy
kahnat	k'äši	Has no sing	Has no singular	Has no singular Definite	?ìtom kahnat	kahnat
clothes	garments such as trousers that are worn to cover the body	Has no sing	Has no indefinite	Has no singular definite	the clothes	clothes
kìdan	säbat zìgäbruwo zìtäsäfäyä Aaläba	kìdan	kìdan	?ìtu kìdan	?ìtom kìdawnti	kìdawnti
earnings	money which is earned by working	Has no sing	Has no indefinite	Has no singular definite	the earnings	earnings
?ìtot	nay dìxam nay sìraH waga	?ìtot	?ìtot	?ìtu ?ìtot	Has no plural definite	Has no plural
father	any male animal in relation to its young	father	a father	the father	the fathers	fathers
?abbo	täbaAtay wäladi	?abbo	?abbo	?ìtu ?abbo	?ìtom ?abbotat	?abbotat
fathers	any of the people from whom somebody is descended	Has no sing	Has no indefinite	Has no singular definite	Has no Plural Definite	fathers
wälädi	Haräg wäläd̈o	Has no sing	Has no singular	Has no singular Definite	?ìtom wälädi	wälädi
fish	a creature that lives in water	fish	a fish	the fish	the fishes	fishes
Aasa	?ab baHri zìnbìr	Aasa	Aasa	?ìtu Aasa, ?ìta Aasa	?ìtom Aasatat, ?ìtän Aasatat	Aasatat
foot	movable part of the body on which a person or animal stand	foot	a foot	the foot	the feet	feet
?ìgri	säb konä ?ìnsìsa zìxädälun t'ät'äw zìblälun kìfili ?akal	?ìgri	?ìgri	?ìtu ?ìgri, ?ìta ?ìgri	?ìtom ?a?gar, ?ìtän ?a?gar	?a?gar
fox	a small dog-like flesh-eating wild animal	fox	a fox	the fox	the foxes	foxes
wäxarya	kìndi kälbi ?ìtìxäwìn t'eläbägiA ?ìtìbälìA ?arawit	wäxarya	wäxarya	?ìtu wäxarya, ?ìta wäxary	?ìtom wäxaru, ?ìtän wäxaru	wäxaru
girl	a young woman who is not married	girl	a girl	the girl	the girls	girls
gwal	zäytämärAawät dìngìl	gwal	gwal	?ìta gwal	?ìtän ?agwalat	?agwalat
knife	a blade fixed in a handle used for cutting as a tool or weapo	knife	a knife	the knife	the knives	knives
karra	nìmäHìrädi mäxtäfi zägälgìl bìHas'in zìsìraH ?ax'Ha	karra	karra	?ìtu karra, ?ìta karra	?ìtom kararu, ?ìtän kararu	kararu
lady	a woman with dignity and good manners	lady	a lady	the lady	the ladies	ladies
wäyzäro	nìzìtämärAwät säbäyti zìwähab mäAarìg	wäyzäro	wäyzäro	?ìta wäyzäro	?ìtän wäyzäraz̀ìr	wäyzäraz̀ìr
leaf	green flat parts of a plant growing from a stem or branch	leaf	a leaf	the leaf	the leaves	leaves
k'os'li	?ab ?a?wam zìbäk'ul k'ät'äliya Hibri zäläwo ?akal	k'os'li	k'os'li	?ìtu k'os'li, ?ìta k'os'li	?ìtom k'os'litat, ?ìtän k'os'litat	k'os'litat

Headword	ConExp	Singular	Indefinite	S-Definite	P-Definite	Plural
life	the active force in animals and plants	life	Has no indefinite	the life	the lives	lives
hìywät	säb ?ìnsäsa täxli wäzätä mìnbar zìx'ì?ilulu Hayli	hìywät	hìywät	Has no singular Definite	Has no plural definite	Has no plural
loaf	a mass of bread shaped and baked in one piece	loaf	a loaf	the loaf	the loaves	loaves
Hìnbaša	bìHìrus' sìrnay zìsìnkät rìgud zìbälä mägbi	Hìnbaša	Hìnbaša	?itu Hìnbaša, ?ita Hìnbaš	?ìtom Hanabìš, ?ìtän Hanabìš	Hanabìš
louse	small wingless insect that live on the skin and hair	louse	a louse	the louse	the lice	lice
x'umal	däm säb ?ìndamäs'äyä zìnäbìr balìA	x'umal	x'umal	?itu x'umal	?ìtom x'umalat	x'umalat
man	the human race	man	Has no indefinite	Has no singular definite	Has no Plural Definite	Has no plural
säb	zär?i säb zäbälä fit'ur kulu	säb	säb	Has no singular Definite	Has no plural definite	Has no plural
man	an adult male human being	man	a man	the man	the men	men
säb?ay	bäs'Hi wädi täbaAtay	säb?ay	säb?ay	?itu säb?ay	?ìtom säb?ut	säb?ut
manner	a personal way of acting or behaving towards other people	manner	a manner	the manner	Has no Plural Definite	Has no plural
t'äbay	s'ìbux' yìxun Hìmax' bahrìy säb zìs'ìwäAalu kal	t'äbay	t'äbay	?ìtu t'äbay	Has no plural definite	Has no plural
men	an adult male human being	man	a man	the man	the men	men
säb?ut	bäs'Hi wädi täbaAtay	säb?ay	säb?ay	?itu säb?ay	?ìtom säb?ut	säb?ut
mother	any female animal in relation to its young	mother	a mother	the mother	the mothers	mothers
?ìnno	wäladit	?ìnno	?ìnno	?ìta ?ìnno	?ìtän ?ìnnotat	?ìnnotat
news	fresh information or reports of recent events	news	Has no indefinite	the news	Has no Plural Definite	Has no plural
zena	?iwanawi wärä	zena	zena	?ìtu zena, ?ìta zena	Has no plural definite	Has no plural
objective	an aim	objective	an objective	the objective	the objectives	objectives
Aìlama	kìtìbäs'ìHo ?ìtìs'ìArälu nägär	Aìlama	Aìlama	?ìtu Aìlama, ?ìta Aìlama	?ìtom Aìlamatat, ?ìtän Aìlamatat	Aìlamatat
ox	a fully-grown male of the cattle family	ox	an ox	the ox	the oxen	oxen
bìAray	nìmaHräs zìAaxälä täbaAtay käbti	bìAray	bìAray	?ìtu bìAray	?ìtom ?abAur	?abAur
patch	small piece of material used to cover a hole or damaged pla	patch	a patch	the patch	the patches	patches
mälgìba	zìtäx'ädädä kìdan ?ìnìlägbälu	mälgìba	mälgìba	?ìtu mälgìba, ?ìta mälgìba	?ìtom mälgìbatat, ?ìtän mälgìbatat	mälgìbatat
people	persons in general	Has no sing	Has no indefinite	Has no singular definite	the people	people
hìzbi	bìzHi zäläwom säbat	hìzbi	hìzbi	?ìtu hìzbi	?ìtom hìzbitat	hìzbitat
prisoner	a person or animal held with limited freedom of movement	prisoner	a prisoner	the prisoner	the prisoners	prisoners
?ìsur	zìtätHazä säb	?ìsur	?ìsur	?ìtu ?ìsur, ?ìta ?ìsur	?ìtom ?ìsurat, ?ìtän ?ìsurat	?ìsurat

Headword	ConExp	Singular	Indefinite	S-Definite	P-Definite	Plural
shelf	a flat long and narrow board for putting or storing things on	shelf	a shelf	the shelf	the shelves	shelves
käbHi	?ax'Hu mäs'aHìfti ?ìnäx'ämìt'älu	käbHi	käbHi	?ìtu käbHi	?ìtom käbHitat	käbHitat
star	distant bodies appearing as a point of light in the sky at nigh	star	a star	the star	the stars	stars
koxob	nay gäza? rì?su bìrhan zäläwo ?ab t'äfär zìrkäb	koxob	koxob	?ìtu koxob, ?ìta koxob	?ìtom käwaxìbti, ?ìtän käwaxìbti	käwaxìbti
team	group of people who work or play together	team	a team	the team	the team	team
ganta	?ab zìxonä s'äwäta bìHadä xoynom mìskal?ot zìgät'mu säb	ganta	ganta	?ìta ganta	?ìtän gantatat	gantatat
thief	a person who steals	thief	a thief	the thief	the thieves	thieves
leba	zäynatu zìwäsìd	leba	leba	?ìtu leba, ?ìta leba	?ìtom lebatat, ?ìtän lebatat	lebatat
tooth	small hard bony objects growing inside the mouth	tooth	a tooth	the tooth	the teeth	teeth
sìnni	mìs gìrs'an zìtätäHaHazä nìmägbi mäHyäxi zìtäk'ìm	sìnni	sìnni	?ìtu sìnni, ?ìta sìnni	?ìtom ?asnan, ?ìtän ?asnan	?asnan
troop	group of soldiers	troop	a troop	the troop	the troops	troops
särawit	bìzìHi zäläwo nìx'inat zìtädaläwä ?as'war zìHazä säb	särawit	särawit	?ìtu särawit	Has no plural definite	Has no plural
water	a liquid that falls as rain, is in lakes, rivers, seas	water	Has no indefinite	the water	Has no Plural Definite	Has no plural
may	?ab Aila baHri k'älay zìrkäb Hìbrìn mäk'ärätìn zäybìlu zìsìtä	may	may	?ìtu may	Has no plural definite	Has no plural
wife	the woman to whom a man is married	wife	a wife	the wife	the wives	wives
säbäyti	bäAälti säb?ay	säbäyti	säbäyti	?ìta säbäyti	?ìtän ?anìšti	?anìšti
wolf	a wild animal of the dog family which hants other animals	wolf	a wolf	the wolf	the wolves	wolves
täxula	zär?i kälbi zìxonä ?ab dur zìnäbìr tägada?i ?arawit	täxula	täxula	?ìtu täxula, ?ìta täxula	?ìtom täxulatat, ?ìtän täxulatat	täxulatat
youth	a young man	youth	a youth	the youth	the youths	youths
män?ìsäy	Aìdmi?u bìzuH zäykonä säb	män?ìsäy	män?ìsäy	?ìtu män?ìsäy	?ìtom män?ìsäyat	män?ìsäyat

Headword	ConExp	Singular	Indefinite	S-Definite	P-Definite	Plural
shelf	a flat long and narrow board for putting or storing things on	shelf	a shelf	the shelf	the shelves	shelves
käbHi	?ax'Hu mäs'aHifti ?ìnäx'ämìt'älu	käbHi	käbHi	?ìtu käbHi	?ìtom käbHitat	käbHitat
star	distant bodies appearing as a point of light in the sky at nigh	star	a star	the star	the stars	stars
koxob	nay gäza? rì?su bìrhan zäläwo ?ab t'äfär zìrkäb	koxob	koxob	?ìtu koxob, ?ìta koxob	?ìtom käwaxìbti, ?ìtän käwaxìbti	käwaxìbti
team	group of people who work or play together	team	a team	the team	the team	team
ganta	?ab zìxonä s'äwäta bìHadä xoynom mìskal?ot zìgät'mu säb	ganta	ganta	?ìta ganta	?ìtän gantatat	gantatat
thief	a person who steals	thief	a thief	the thief	the thieves	thieves
leba	zäynatu zìwäsìd	leba	leba	?ìtu leba, ?ìta leba	?ìtom lebatat, ?ìtän lebatat	lebatat
tooth	small hard bony objects growing inside the mouth	tooth	a tooth	the tooth	the teeth	teeth
sìnni	mìs gìrs'an zìtätäHaHazä nìmägbi mäHyäxi zìtäk'ìm	sìnni	sìnni	?ìtu sìnni, ?ìta sìnni	?ìtom ?asnan, ?ìtän ?asnan	?asnan
troop	group of soldiers	troop	a troop	the troop	the troops	troops
särawit	bìzìHi zäläwo nìx'inat zìtädaläwä ?as'war zìHazä säb	särawit	särawit	?ìtu särawit	Has no plural definite	Has no plural
water	a liquid that falls as rain, is in lakes, rivers, seas	water	Has no indefinite	the water	Has no Plural Definite	Has no plural
may	?ab Aila baHri k'älay zìrkäb Hìbrìn mäk'ärätìn zäybìlu zìsìtä	may	may	?ìtu may	Has no plural definite	Has no plural
wife	the woman to whom a man is married	wife	a wife	the wife	the wives	wives
säbäyti	bäAälti säb?ay	säbäyti	säbäyti	?ìta säbäyti	?ìtän ?anìšti	?anìšti
wolf	a wild animal of the dog family which hants other animals	wolf	a wolf	the wolf	the wolves	wolves
täxula	zär?i kälbi zìxonä ?ab dur zìnäbìr tägada?i ?arawit	täxula	täxula	?ìtu täxula, ?ìta täxula	?ìtom täxulatat, ?ìtän täxulatat	täxulatat
youth	a young man	youth	a youth	the youth	the youths	youths
män?ìsäy	Aìdmi?u bìzuH zäykonä säb	män?ìsäy	män?ìsäy	?ìtu män?ìsäy	?ìtom män?ìsäyat	män?ìsäyat

DECLARATION

THIS THESIS IS MY ORIGINAL WORK AND HAS NOT BEEN SUBMITTED FOR DEGREE IN ANY OTHER UNIVERSITY

Mulu Gebreegziabher

THE THESIS HAS BEEN SUBMITTED FOR EXAMINATION WITH MY APPROVAL AS UNIVERSITY ADVISOR

DR. GEBREMEDHIN SIMON

ATO WORKSHET LAMENEW