



ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULTY OF INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

**DESIGN AND IMPLEMENTATION OF MULTILANGUAGE
ELECTRONIC DICTIONARY FOR SMART PHONES:**

**A DICTIONARY OF AMHARIC, AFAAN OROMO, ENGLISH AND
TIGRIGNA LANGUAGES**

By

Amsale Zelalem

A Project paper submitted to the School of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science

June 14, 2010

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
FACULTY OF INFORMATICS
DEPARTMENT OF COMPUTER SCIENCE

**DESIGN AND IMPLEMENTATION OF MULTILANGUAGE
ELECTRONIC DICTIONARY FOR SMARTPHONES:**

**A DICTIONARY OF AMHARIC, AFAAN OROMO, ENGLISH AND
TIGRIGNA LANGUAGES**

By:

Amsale Zelalem

Advisor:

Solomon Atnafu (PhD)

Examining Board:

1. Dr. Solomon Atnafu, Advisor _____
2. _____
3. _____

Table of Contents

1. Introduction	1
1.1. Overview	1
1.2. Statement of the problem	4
1.3. Motivation	4
1.4. Objective	5
1.5. Scope and Limitation	6
1.6. Methodology and Procedures	6
1.7. Organization of the document	7
2. Overview of Electronic Dictionary and the Languages.....	8
2.1. Overview of Electronic Dictionary	8
2.2. Overview of the Amharic, Afaan Oromo and Tigrigna Languages.....	12
2.3. Electronic Dictionary for Ethiopian Languages and the Challenge.....	18
3. Related Work.....	20
3.1. Monolingual ED.....	20
3.2. Multilingual ED.....	21
3.3. Mobile ED's	23
4. Requirement Analysis.....	25
4.1. Proposed system.....	25
4.1.1. Overview.....	25
4.1.2. Functional Requirements	28
4.1.3. Non-Functional Requirements	28

4.1.4.	System Model	28
5.	System Design	39
5.1.	Design Goals	39
5.1.1.	Performance criteria.....	39
5.1.2.	End User Criteria	40
5.1.3.	Maintenance Criteria.....	41
5.1.4.	Dependability Criteria.....	41
5.2.	Architecture of the Proposed System	41
5.3.	Subsystem Decomposition	43
5.4.	Persistent Data Management	45
6.	Implementation of the Multilingual Dictionary.....	50
6.1.	Development Environment	50
6.2.	Development Tools	51
6.3.	The Ethiopic Multilingual Dictionary Application	53
7.	Conclusion and Recommendation	61
7.1.	Conclusion.....	61
7.2.	Recommendation.....	62
	References.....	64

List of Figures

<i>Figure 2.1 Two-step technical-lexicographic electronic-dictionary typology</i>	11
<i>Figure 4.1 Work flow diagram of the mobile dictionary system</i>	27
<i>Figure 4.2 Use Case diagram of the proposed system</i>	29
<i>Figure 4.3 Sequence diagram for Choose Dictionary Type use case</i>	33
<i>Figure 4.4 Sequence diagram for Get Suggestion use case</i>	34
<i>Figure 4.5 Sequence diagram of Switch Type use case</i>	35
<i>Figure 4.6 Sequence diagram for Lookup Match use case</i>	36
<i>Figure 4.7 Activity diagram of the system</i>	37
<i>4.8 Class diagram of the system</i>	38
<i>Figure 5.1 System architecture of the Ethiopic multilingual dictionary system</i>	42
<i>5.2 Subsystem decomposition of the Ethiopic multilingual dictionary system</i>	43
<i>5.4 Table designs of the four languages</i>	47
<i>Figure 6.1 Android 2.0.1 platform Emulator</i>	52
<i>Figure 6.2 Eclipse IDE with the ADT plug-in-installed</i>	53
<i>Figure 6.3 The Ethiopic Multilanguage electronic dictionary icon</i>	54
<i>Figure 6.4 The startup screen for the Ethiopic Multilanguage dictionary application</i>	55
<i>Figure 6.5 The list of dictionaries included in the multilingual dictionary</i>	55
<i>Figure 6.6 Screenshot for the Tigrigna Dictionaryentry subtype</i>	56
<i>Figure 6.7 List of soft keypads available on the device</i>	57
<i>Figure 6.8 user selected keypad displayed on activity</i>	58
<i>Figure 6.9 suggested word list displayed</i>	59
<i>Figure 6.10 Translation and pronunciation of the Tigrigna word “ሓመ-ሸተ” in Afaan Oromo</i>	60

List of Tables

<i>Table 2.1 Amharic Fidels and their Sound</i>	13
<i>Table 2.2 Amharic sounds with Special script Symbols</i>	14
<i>Table 2.3 Tigrigna Fidels</i>	17
<i>Table 4.1 Choose Dictionary Type use case description</i>	30
<i>Table 4.2 Get Suggestion Use Case Description</i>	30
<i>Table 4.3 Switch Type Use Case Description</i>	31
<i>Table 4.4 Look-Up Match Use Case description</i>	32
<i>Table 5.1 algorithm for displying list of suggested words</i>	48

List of Acronyms

ADT	Android Development Tools
AVD	Android Virtual Device
API	Application Programming Interface
CPU	Central Processing Unit
CV	Consonant-Vowel
DVM	Delvik Vertual Machine
ED	Electronic Dictionary
GUI	Graphical User Interface
IDE	Integrated Development Environment
NLP	Natural Language Processing
OHA	Open Handset Alliance
OS	Operating System
PC	Personal Computer
PDA	Personal Digital Assistant
PD	Paper Dictionary
SD	Secure Digital
SDK	Software Development Kit
SQL	Structured Query Language
UI	User Interface
UML	Unified Modeling Language
XML	Extensible Markup Language

Abstract

The ever increasing growth of the mobile technology is giving birth to more powerful devices like smart phones which provide many of the functionalities of personal computers in addition to making phone calls. As these devices are becoming more and more functional, they are penetrating swiftly into every society of the world. As a result, there is a need to design and develop language specific applications that can help users interact with these devices through the language they understand. Electronic Dictionaries (EDs) are one example of such applications which can be used as an electronic reference resource containing a library of words and their meanings. As globalization is bringing the world into one, understanding several languages is very crucial for smooth interaction and wide communication with others. Thus, several Multilanguage dictionary applications with two or more languages are becoming integrated on handheld devices like PDAs and smart phones to cope up with the globalization issue. Having such Multilanguage dictionary application on mobile devices is very helpful as it makes language learning easier and faster, can be accessed anytime and anywhere needed, no need of extra device purchasing and the like.

However, the currently available dictionary applications on these devices cover mostly the western languages and no focus had been given to develop these applications for languages like the ones spoken in Ethiopia. In countries like Ethiopia, with various language sets, such a Multilanguage dictionary is essential for ease of communication among citizens of the country as well as other foreigners.

This report presents the first attempt for design and implementation of a Multilanguage ED for Amharic, English, Afaan Oromo and Tigrigna languages on handheld devices. The system has been successfully developed and tested on an emulator of an Android based smart phones before deploying it on the target device. The Android platform is chosen among other mobile platforms as it is designed to support different script types including Ethiopic scripts, and has a promising future to be used on all handheld devices. Finally, the system has been tested on Android based HTC Hero smart phone.

Keywords: Electronic Dictionary, Multilingual Electronic Dictionary, Dictionary on Smart phone, Multilingual Dictionary for Ethiopian Languages

Acknowledgment

First and foremost, thank you God for making me who I am today.

I would like to express my heartfelt thanks to my advisor Dr. Solomon Atnafu without whom this project work would not come into existence. He provided me with the relevant guides, advises and encouragements that I needed throughout the project.

I also take this opportunity to express my profound gratefulness to my families for always believing in me and for their unstopped support and patience during my study.

I am thankful to Andualem Workneh for giving me valuable technical assistances, Workagegnehu for his laborious work of revising my document and providing me important comments in this project. Dear Fetiya, I thank you very much for your undisputable support since we became friends.

Finally, all my friends and staffs whom I did not list out here also deserves my gratitude as you all encouraged me, suggested me some ways to improve this work and covered my duty on those busiest days of my work.

Chapter One

Introduction

1.1. Overview

An advance in the communication and computer technology is enabling the world to access different kinds of information instantly. Devices are becoming smaller and cheaper while they are more user friendly and powerful in carrying user oriented data. Personal Digital Assistants (PDA) and cell phones are the major recent outcomes of such advancements where the first one is mainly used for data storage having only a few communication features while the later one is used mainly for phone communication [1].

In recent years, many manufacturers have begun to blur the line between PDA and cell phone. A smart phone has been introduced to the market that incorporates the two devices features together and provides more advanced applications in one. A smart phone is a mobile phone offering advanced capabilities, often with PC-like functionality. As a whole, a smart phone is a convergence of the “3-C”: Communication, Computing and Consumer Electronics [1].

Smart phones have been regarded as the most important and influential personal communication equipment ever developed in the history of mankind [2]. Presently, they are being used in every aspect of business along with every facet of personal life. Smart phones and, mobile phones in general, have gained widespread acceptance due to their portability, versatility, and inexpensive price in both developed and developing countries. In Ethiopia, as one part of the world, the use of such devices is becoming popular especially among youths.

The first smart phone was designed by IBM in 1992 [3] and since then a lot of companies are involved in designing and releasing the different kinds of currently available smart phones throughout the world. New advanced handsets and the popularity of mobile applications contributed to increased smart phone buzz in 2009. A report by [4] shows despite flat handset sales over the past year, smart phone shipments rose 24% worldwide. The smart phones can be broadly categorized into two based on the platform they are using as proprietary and open source. Those smart phones with Microsoft, Palm and BREW (Binary Runtime Environment for Wireless) Operating System can be named as proprietary since the operating systems (OS) they are using is proprietary and provide largely restricted application programming interfaces (APIs)

while those with the Android platform are referred to be open source as the OS they use is an open source with unrestricted API that allows developers to design and incorporate different applications on the smart phones.

Android is a popular open source software platform for mobile devices based on the Linux operating system which had been developed by Google together with the Open Handset Alliance (OHA) and released in 2007[5]. In addition to the typical APIs for screen drawing, user input, and network access, Android provides APIs for storage, media, graphics, and even direct hardware access. Media APIs are available for both playback and recording of audio, video, and still images. For storage, developers aren't limited to file-based APIs. SQLite is available for relational data storage, a preferences API is available for simple setting storage, and applications can extend the data storage mechanisms available [6]. Moreover, the Android Operating System also provides a convenient environment for developing localized applications as it supports different natural languages. As a result, mobile application developers around the world are taking advantage of this platform and developing several powerful language specific applications like dictionaries for Android based smart phones.

Recently, full-fledged dictionary software and complete suites of dictionary databases for several languages are becoming available on the smart phones and PDAs [7]. A dictionary can be defined as a book that represents most commonly spoken words of a language together with their pronunciations and definitions. It can contain the word's definitions using the same language the word is written at, in which case it is known as a monolingual dictionary, or in two or more different languages which is then termed as bilingual or multilingual dictionary respectively.

In its electronic form, a dictionary can be defined as a textual database to be used in any natural language processing system. An electronic dictionary is either a small handheld computer with integrated reference materials, or a PDA or a smart phone with a dictionary program. Electronic dictionaries (EDs) are also programs that can be downloaded from the Internet or purchased on a CD-ROM or DVD and installed on a desktop computer or on a laptop [7]. The major differences between EDs and paper dictionaries (PDs) are size, weight and cost [8]. Moreover, an ED can also be modified much more frequently as there is no need for a new edition of the dictionary when a few new words come into being or when the meanings of some words change in a few subtle ways. The size and the contents of EDs can vary a lot according to the target of the

attended public, the target device on which they are going to be deployed and the cost of the collected resources. The main advantage of having dictionaries on handheld devices like smart phones is that they are lightweight and compact, can be accessed at anytime and anywhere needed, and there is no need for extra-device to be purchased.

Several world languages provide dictionaries to be used as a reference book by language learners. In Ethiopia also, dictionary books of some local and international languages like Amharic-English are available on market.

Amharic, Oromiffa and Tigrigna are the three most widely and commonly used languages in Ethiopia and English has been the most spoken and used foreign language in many non-English speaking countries including Ethiopia. There are some paper dictionaries published for Amharic, Oromiffa as well as Tigrigna languages and also other languages used in Ethiopia. These dictionaries can be used by foreign tourists as well as the country's citizens to learn about those languages they are interested in. Most importantly, the dictionaries can foster the tourism sector greatly as they provide common and basic words of some languages spoken in the country which lets the tourists interact with the speakers of that language and enjoy their stay. Moreover, they can also have some contribution towards strengthening globalization as they let the other world know about some of the languages spoken in the country. On top of these, the dictionaries can help the people of the country who are found in different regions to know the languages of others in another region and communicate with each other using languages different from their mother tongue. As a whole, these dictionaries can be used as the best learning tool for the languages they encompass.

Language learning involves memorization and practice of a large number of vocabulary words and grammatical structures. Any language learning can be divided into four: listening, reading, speaking, and writing skills out of which reading and vocabulary knowledge are two of the most important components of performance in a second language and depend on the other [9]. Language learning has been one stream where mobile technology in recent times, has been popular by providing instant help on obtaining word's definition and pronunciation [10].

As the number of mobile device users in Ethiopia is increasing from day to day, developing mobile dictionary applications for effective learning of the widely spoken languages together

with English language can be even more advantageous to fulfill local users need and those who need to use the languages, in general.

1.2. Statement of the problem

As globalization is bringing the whole world into one, newly invented technological devices such as smart phones, are able to penetrate into any society with an amazing speed. Today's digital technology is being designed by paying more attention to western languages like English. But this has a great influence on other language speakers, like us, who use the devices extensively but could not get any application written for their language.

In Ethiopia, the effort exerted on developing local applications for these newly emerging devices like smart phones is in its infant stage where some students try to develop third-party applications as a final year project but none exist on the market. Having such applications on mobile devices have significant contribution to use the technology more effectively in the local context. As a result, designing and implementing more and more localized applications like a multilingual dictionary of some local languages on the smart phones is essential for making the devices extensible and more usable by the local people.

Moreover, even though a dictionary book has several advantages like the ones mentioned above, it is not exhaustively used as it is supposed to. The reason behind this limited usage is that first the books are not abundantly found in an electronic form which would have been more preferable and easy way of storing and accessing them especially by literates. Secondly, it is accompanied by extra cost as it is difficult to get a four-in-one dictionary book that holds the needed languages in one. This forces the user to buy and have several books of the languages under interest which also makes the user to carry cumbersome paperback dictionaries wherever she/he goes.

1.3. Motivation

In this era of advanced technology, the way people used to communicate, work and live have been changed a lot as new and improved choice of communication is offered. These choices of communication, in turn, provide a potential solution to meet some challenges like demand for more flexibility in delivery of information in terms of time, location, content, and form. People

are taking benefit of such technology and start representing and exchanging data electronically through mobile devices.

There seems to be a constant stream of new technology breaking into the mobile phone market. The huge expansion of these devices with large storage capacity, open platform and that are capable of integrating several applications also allows developers to add more user oriented functionalities. As a result, a useful possibility would be to integrate and access different local applications from such devices.

Moreover, EDs of several world languages are becoming available in almost all smart phones and are being used as a language learning tool [7]. But there does not exist one for any of the Ethiopian languages. Ethiopia is a multi-linguistic country with eighty or so speech varieties and ethnic groups. It is crucial that all the languages obtain appropriate attention, are refined and promoted empowering them to better serve the speakers. Essential aspects of the life of a people are encoded and expressed in its language. It is, among other things, this inner complexity of a society that is captured in studying the languages [11]. Hence, integrating an electronic dictionary of some of the local languages would move the present teaching-learning mechanism of the languages one step forward.

1.4. Objective

General objective

The general objective of this project is to design and implement a four-in-one Multilanguage ED comprising of Amharic, English, Afaan Oromo and Tigrigna languages for Android based smart phones using the Ethiopic and Latin character sets.

Specific objectives

The specific objective includes:

- Design a suitable user interface for smart phones that allows the user to interact with the phone through the language he/she understood
- Construct a database that stores words and their corresponding definitions and pronunciations for each language
- Enable the user to write words with the language he/she knows

- Design keyword searching scheme
- Retrieve words, their definitions and pronunciations from the database
- Implement the designed system and test it on real devices

1.5. Scope and Limitation

Though Ethiopia is a country with several ethnic groups living together and each having their own language, this project only targets on developing a multilingual dictionary application for the three languages of the country, as they are the ones spoken by most, and English language is also taken from all other foreign languages as it is assumed most foreigners can speak and understand the language and is also the main working International language in the country. However, it might be possible to incorporate other languages as well. Moreover, this application is designed and developed only for smart phones with Android platform.

1.6. Methodology and Procedures

For the successful completion of this project, a number of analyses, programming tools and techniques are put into use including:

- Perform literature review thoroughly and assess related works that focus on developing a multi-language dictionary system.
- Study the natures and behaviors of the Amharic, Afaan Oromo and Tigrigna languages.
- Identifying system requirements. This includes identifying the appropriate data sources to be used as well as identifying functional and non-functional requirements of the application.
- Designing the system based on the identified requirements.
- Actual implementation and development of the system using the chosen programming languages and development tools.
- Testing and validating the developed system according to the identified requirements

1.7. Organization of the document

The rest of this document is organized as follows. Chapter two provides a general overview about Electronic Dictionary and the languages chosen for this project. Related works regarding design and implementation of Electronic Dictionary are discussed in chapter three. Chapter four and five present the requirement analysis and system design of the developed system respectively. In the remaining chapters, implementation of the system, conclusion and future works are briefly explained.

Chapter Two

Overview of Electronic Dictionary and the Languages

2.1. Overview of Electronic Dictionary

We all encounter very often the problem of not being able to find or remember the word expressing the idea we have in our mind. If we have the time we may reach for a paper dictionary. Yet, this kind of resource may be of little help, if it does not contain the word we are looking for or it expects from us precisely what we are looking for. The better solution would be to construct an ED that imitates the various methods people use to search for words and definitions in their mind [12].

A dictionary is a reference book that intends on providing words and phrases definitions, including multiple meanings as the word is used in a language. The most frequently used dictionary is a language dictionary that includes the most frequently used words in a language [13]. Language dictionaries are useful for all and are made for different types of users: scholars, office workers, schools, second language learners, etc.

Dictionaries can be classified into three as monolingual, bilingual and multilingual based on the number of languages they contain. A dictionary of words and their definitions using the same language the word is written at is known as a monolingual dictionary. Such kinds of dictionaries are helpful in elaborating a given word in detail by providing usage examples and the like. The bilingual dictionaries contain words in one language and their definition in another. When the translation languages included are more than one, then the dictionaries are termed as multilingual. The dictionaries, monolingual, bilingual, or multilingual, are the standard way of collecting and presenting lexicographic knowledge about one or more languages [37].

There are several monolingual as well as bilingual paper dictionaries being published by different authors for some of the Ethiopian languages. The Concise Amharic-English Dictionary, published by Amazon is the first bilingual and bidirectional dictionary book for Amharic language together with the English language having phonetic transcriptions [14]. The Amsalu Aklilu Amharic-English dictionary [15] is also one of the most widely known and used bilingual paper dictionary which constitutes large number of commonly used Amharic words but it is not a

bidirectional dictionary that is it only has Amharic words with English definitions and not vice versa.

Some efforts had also been observed to publish dictionaries of some other languages used in Ethiopia like Oromiffa and Tigrigna but are not as abundant as the Amharic ones. Oromo-English Dictionary, by [16] and Oromo Technical Terms dictionary by [17] are some of the significant paper dictionaries that had been published for Afaan Oromo language.

In addition to the monolingual as well as bilingual paper dictionaries, some multilingual dictionaries are also becoming available which consists of Oromiffa-Amharic-English [48] and Amharic-English-Tigrigna languages. These dictionaries are more advantageous than the monolingual as well as the bilingual dictionaries in that they provide a three-in-one compiled package and are lighter to carry as compared to having two or three separate dictionaries and also reduce the cost incurred while purchasing several books. However, the currently available multilingual dictionaries of these languages have limitation which is lack pronunciations of each translated word for each language which is very helpful for those who do not know how to read the other languages. For instance, a person who is familiar with the English letters, can read text written with the Latin scripts and vice versa but could not be able to read those texts written in some Ethiopian languages like Amharic that uses the Ethiopic scripts. Hence there needs to be a way for the person to read and understand those Ethiopic words with the one he/she is familiar with so that the user can feel free to communicate with the native people. This can be achieved by providing a pronunciation of the translated words of the other languages using the users' language. Moreover, as of our knowledge, there exist no published dictionary book consisting of the three widely spoken languages in the country; namely Amharic, Oromiffa and Tigrigna altogether which is a very important and a must to have dictionary for the successful learning of these languages.

These local dictionaries can be used by students, foreign tourists as well as literates in the country to learn and understand about those languages they are not familiar with. The dictionaries also have a great contribution towards preserving words of a given language and introducing it to one another of the language speakers. The availability of these dictionaries can also create a mutual and suitable environment to exchange information and communicate with each other easily among the citizens of the country who speaks different language. In general

they can be used as learning aid by providing words of one language translated into other languages. Not only students, local citizens and tourists, but others can also benefit from using these dictionaries as a learning aid. Employees of international organizations and companies, diplomats and advisors from other countries, geologists, etc. that continuously work in Ethiopia requires language training and a multilingual dictionary would be the best teacher for them. The Diasporas living abroad can also use these dictionaries to teach their kids about their mother tongue and pass the language to the next generation. Foreigners who adopt Ethiopian children can also benefit from using these books as they can help them communicate the adopted children with their mother tongue and at the same time teaching them their own language.

In its electronic form, a dictionary can be defined as an electronic reference resource that contains collection of words with their meanings, spellings, and etymologies [18]. Mostly, EDs have similar function to paper dictionaries in that they are searchable and they allow one to find specific fragments of information about words. But sometimes they also work in the background of other programs, such as word processors, as spell checkers, to ensure proper spelling of a given word.

A two step technical ED typology had been proposed by Lehr [19] as shown in figure 2.1 below. The first step classifies the EDs on technical grounds, the main dichotomy being online vs. offline dictionaries, the offline dictionaries being further divided into pocket electronic dictionaries and PC dictionaries. In a second step, each of these EDs can then be evaluated on lexicographic grounds. Either an ED is based on a paper dictionary or it is a new development.

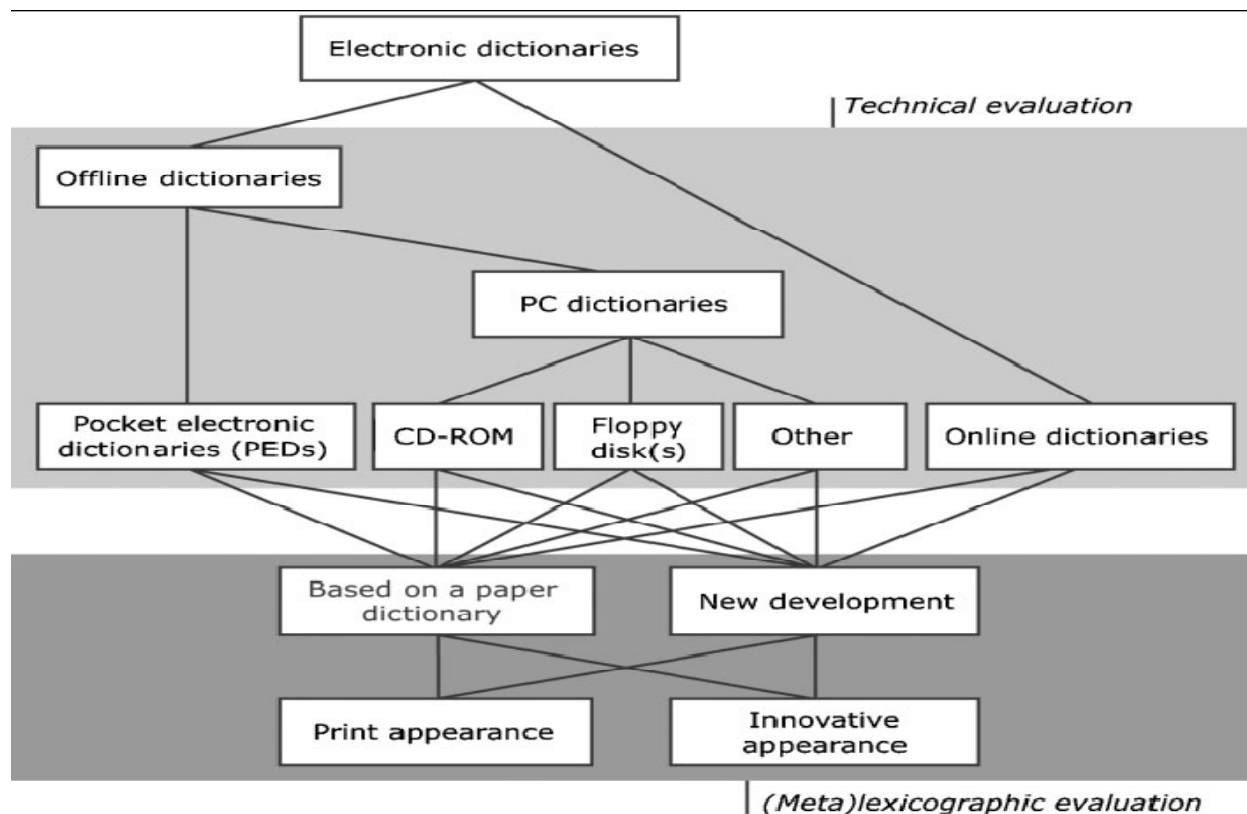


Figure 2.1 Two-step technical-lexicographic electronic-dictionary typology [19]

One can find EDs for several world languages on many different devices and in many places on the Internet which can be accessed by anyone for free. Some portable, battery-operated devices are made exclusively to serve as electronic dictionaries. Many mobile devices like smart phones and PDAs also contain integrated electronic dictionaries that can be used for checking spelling or for reference purposes. The EDs are made available either by the dictionary book publishing companies, by an online community or even by an individual who volunteer to let others use a given dictionary resource. While constructing an ED, one can use freely available lexicon of the language as a source of data or add the words and definitions manually.

Having an electronic version of the above local dictionaries would be crucial as the trend of data representation and information exchange is shifting from paper to electronic form and the communication technology is highly advancing providing smaller, mobile and portable devices with lots of functionalities. Customizing the paper dictionaries to be used in mobile devices would be even much more convenient and essential than the paper dictionaries as it contains many more words in a much smaller space. Also, it tends to be much faster to search an electronic dictionary than to search a paper dictionary; one only needs to type in the desired word

to access all of the information associated with it. It also elevates the problem of finding these dictionaries wherever the user is and whenever needed.

2.2. Overview of the Amharic, Afaan Oromo and Tigrigna Languages

- **The Amharic Language**

Amharic is one of the Semitic languages spoken in most part of Ethiopia. It is the second most-spoken Semitic language in the world, after Arabic [20], and serves as the official language of Ethiopia which makes it to be widely used all over the country. Amharic is also the working language of several regions within the federal system of the country, including the Amhara Region. Outside Ethiopia, around 2.7 million emigrants also use the Amharic language notably those found in Egypt, Israel and Sweden [21].

Amharic is written with a version of the Ge'ez/Ethiopic script known as ፊደል (Fidel) and have rich legacy of both typeset and calligraphic literature [22]. Today researchers from both inside and outside the country are becoming more interested in studying the diverse nature of this language. However, on top of the language's orthography complexity, the unavailability of freely accessible electronic corpora, glossary, and transcription standard are significant barriers to these researchers [22].

The script of Amharic language is phonetic and has 32 consonants and 7 vowels. The orthographic representation of the language is organized into orders making the 32 consonants to have seven different orders (derivatives) reflecting the seven vowel sounds. Out of the seven orders, six are consonant-vowel (CV) combinations while the seventh, which is the sixth order, can be the consonant itself or sometimes it can take the vowel ∂ in a spoken form [23]. Amharic words are written from left-to-right, unlike other Semitic languages and writing a single Amharic Fidel involves combination of one or more English characters as the Fidels are created by CV fusion.

The basic Amharic Fidels together with their derived forms and sounds are shown in Table 2.1 below as taken from [24]

Table 2.1. Amharic Fidels and their sound

Ä	u	i	A	e	ð	o
ሀ hä	ሁ hu	ሂ hi	ሃ ha	ሄ he	ህ hð, h	ሆ ho
ለ lä	ሉ lu	ሊ li	ላ la	ሌ le	ል lð, l	ሎ lo
ሐ hä	ሑ hu	ሒ hi	ሓ ha	ሔ he	ሕ hð, h	ሖ ho
መ mä	ሙ mu	ሚ mi	ማ ma	ሜ me	ም mð, m	ሞ mo
ሠ sä	ሡ su	ሢ si	ሣ sa	ሤ se	ሥ sð, s	ሦ so
ረ rä	ሩ ru	ሪ ri	ራ ra	ራ re	ር rð, r	ሮ ro
ሰ sä	ሱ su	ሲ si	ሳ sa	ሴ se	ስ sð, s	ሶ so
ሸ šä	ሹ šu	ሺ šī	ሻ ša	ሼ še	ሽ šð, š	ሾ šo
ቀ qä	ቁ qu	ቂ qi	ቃ qa	ቄ qe	ቅ qð, q	ቆ qo
በ bä	ቡ bu	ቢ bi	ባ ba	ቤ be	ብ bð, b	ቦ bo
ተ tä	ቱ tu	ቲ ti	ታ ta	ቲ te	ት tð, T	ቶ to
ቸ câ	ቹ çu	ቺ çī	ቻ çā	ቼ çē	ች çð, ç	ቸ ço
ኀ hä	ኁ hu	ኂ hi	ኃ ha	ኄ he	ኅ hð, h	ኆ ho
ነ nä	ኑ nu	ኒ ni	ና na	ኔ ne	ነ nð, n	ኖ no
ኘ nä	ኙ ñu	ኚ ñī	ኛ ñā	ኜ ñē	ኝ ñð, ñ	ኞ ño
አ a	ሁ u	ሀ i	አ a	ሄ e	አ ð	አ o
ከ kä	ከ ku	ከ ki	ካ ka	ከ ke	ክ kð, k	ኮ ko
ኸ hä	ኹ hu	ኺ hi	ኻ ha	ኼ he	ኽ hð, h	ኾ ho
ወ wä	ወ wu	ዐ wi	ዐ wa	ዐ we	ወ wð, w	ዐ wo
ዐ a	ዑ u	ዒ i	ዓ a	ዔ e	ዐ ð	ዑ o
ዘ zä	ዘ zu	ዘ zi	ዘ za	ዘ ze	ዘ zð, z	ዘ zo
ዠ žä	ዡ žu	ዢ žī	ዣ žā	ዤ žē	ዥ žð, ž	ዦ žo
የ yä	የ yu	የ yi	ያ ya	የ ye	የ yð, y	የ yo
ደ dä	ደ du	ደ di	ደ da	ደ de	ደ dð, d	ደ do
ጀ jä	ጁ ju	ጂ ji	ጃ ja	ጄ je	ጅ jð, j	ጆ jo
ገ gä	ገ gu	ገ gi	ገ ga	ገ ge	ገ gð, g	ገ go
ጠ ṭä	ጡ ṭu	ጢ ṭī	ጣ ṭā	ጤ ṭē	ጥ ṭð, ṭ	ጦ ṭo
ጨ çä	ጩ çu	ጪ çī	ጫ çā	ጬ çē	ጭ çð, ç	ጮ ço
ጰ p̄ä	ጱ p̄u	ጲ p̄ī	ጳ p̄ā	ጴ p̄ē	ጵ p̄ð, p̄	ጶ p̄o
ጸ šä	ጹ šu	ጺ šī	ጻ šā	ጼ šē	ጽ šð, š	ጾ šo
ፀ ṣä	ፁ ṣu	ፂ ṣī	ፃ ṣā	ፄ ṣē	ፅ ṣð, ṣ	ፆ ṣo
ፈ fä	ፉ fu	ፊ fi	ፋ fa	ፌ fe	ፍ fð, f	ፎ fo
ፐ pä	ፑ pu	ፒ pi	ፓ pa	ፔ pe	ፕ pð, p	ፖ po
፱ vā	፳ vu	፴ vi	፵ va	፶ ve	፷ vð, v	፸ vo

Among the 32 consonants, Fidel **ፆ** and **ፈ** are considered as semi-vowels.

Regarding Amharic words and statements pronunciation, there is no such internationally agreed way of transliterating them into Roman characters which makes it very difficult for non-speakers of the language to learn and understand it [24]. As can be seen from Table 2.1 above, The Fidels **ፆ** and **ፈ**, for instance, have same sound which is sä and hence they are being used interchangeably as a single consonant even though they have different labels. Like these two Fidels, there are also other with same sound like **ሀ**, **ሐ**, **ኀ** and **ኸ** whose sound is hä.

Some of the Amharic consonant sounds are almost the same as English sounds, but there are also other sounds that could not be found in English. b, p, m, f, w, s (as in ‘sun’), z, y, g (as in ‘go’), k and h have approximately same sound in both languages[24]. There are also some other sounds that are similar in both languages but are differently rendered in the phonetic script by special symbols. These sounds are shown in Table 2.2 below.

Table 2.2. Amharic sounds with special script symbols

Phonetic Script	Corresponds to
ĉ	ch like in church
š	sh like in shine
ž	s like in pleasure
ñ	gn like in Champaign

Those sounds which are found only in Amharic and not in English are transcribed with special symbols like ç and ț.

There is also no exact correspondence between the pronunciation of Amharic and English vowels [24]. The Amharic vowel (i) is pronounced as ‘ee’ in English like in ‘feet’. Thus, the word “ፊት” , for instance, ‘face’ in English is pronounced as ‘fit’. The vowel (e) in Amharic sounds as ‘a’ like in that of ‘state’ for example “ቤት” pronounced as ‘bet’ which means ‘house’ in English. The vowel (ä) is pronounced approximately as ‘uh’, and the vowel (a) as ‘ah’. Vowel (o) is pronounced like the English ‘a’ as in ‘also’ and vowel (u) like the English ‘o’ as in ‘who’. And finally, the vowel (ö) sounds like the ‘e’ vowel in English as in ‘stores’.

Germination, in a language, can be described as lengthening of the consonant to represent different meanings for a given word. In Amharic language, no special symbol is used to indicate word germinations [22] and speakers of the language have no problem with that since the meaning of a given word depends on the context it has been used. However this seems to be a problem for non-speakers of the language to clearly understand and pronounce those Amharic words as there is no special symbol to indicate the germination. For example, the word “አለ” can be read as both “alä” ‘he said’ and “allä” ‘there is’ in English. As a rule, germination only occurs in medial or final position of Amharic words but not at initial position [24].

- **Afaan Oromo Language**

Unlike Amharic, Afaan Oromo(also known to be Oromiffa) is part of the Lowland East Cushitic group within the Cushitic family of the Afro-Asiatic phylum and is one of the major Languages that are widely spoken and used in Ethiopia and to some extent in Kenya too [25]. Currently, it is an official language of the Oromia state, the largest region in Ethiopia and instructional medium for primary and junior secondary schools throughout the region and its administrative zones. Among the forty or so languages found in this Cushitic branch of the Afro-asiatic language family, Afaan Oromo is considered as one of the most extensive and widely used language [26]. It is a common mother tongue for Oromo people, who are the largest ethnic group in Ethiopia, which constitutes around 37.1% of the population according to the 2007 census [27]. In addition to these first language speakers, members of other ethnicities who are in contact with the Oromo people also speak Oromiffa as a second language. In general, within Africa, it is the language with the 4th most speakers, after Arabic, Swahili and Hausa [27].

Regarding the writing system of the language, Qubee (Latin-based alphabet) which is similar to the English letters, had been adopted and become the official script of Afaan Oromo since 1991[26]. There are ten vowels and twenty four native consonant phonemes in this language. The vowels are grouped into two as short vowels and long vowels [28]. The short vowels are: a, e, i, o, u while the long vowels are: aa, ee, ii, oo, uu. There are also Qubees of diphthongs which are produced by combining two consonant Qubee together. These are: ch, dh, ny, ph, sh, dz. The sounds of p, v and z characters are not found in Afaan Oromo language but are used in the language as adopted sounds [28]. Currently, Afaan Oromo is widely used as both written and spoken language in Ethiopia and some neighboring countries, including Kenya and Somalia.

Moreover, a number of literature works, newspapers, magazines, education resources, official documents and religious writings are written and published in Afaan Oromo.

Regarding its pronunciation, a sound segment in Afaan Oromo is produced in two types of duration. The first duration is short lived and represented in writing by a single Qubee. The second duration is twice the first duration and therefore represented by doubling Qubees. If similar consonants are double in a word, the sound should be stressed or emphasized [28]. For example the word kommee, which means ‘come’ in English, has two same consonants in the middle. So when reading this word more stress is given to the m character in opposite to the word komee which has single m character and its sound is lighter. However, this repetition or germination of consonants is allowed only in the medial part of a syllable. Moreover, if similar vowels are repeated in a word, the sound should be longer. For instance the word laagaa is more extended than the word laga when read in Afaan Oromo language. When the single vowel ‘a’ occurs at the end of a word, it is pronounced as schwa (inverted e) such as in ‘Uummata’ which means ‘people’ in English whereas it is pronounced (delta) elsewhere. Except for some, most of the letters in the Oromo language have same pronunciation and sound with English.

- **The Tigrigna Language**

Tigrigna is a Semitic language, just like the Amharic language, spoken by the people in the Tigray region of Ethiopia and in central Eritrea. It is also the most spoken language by large immigrant communities around the world mainly in Sudan, Saudi Arabia, the United States, Germany, Italy, the United Kingdom, Canada and Sweden [29]. It has a strong relationship with the Ge’ez language, the oldest language spoken and used in Ethiopia. As of its writing, the Tigrigna language also uses the Ethiopic script but with some extended features from the one used in the Amharic language writing and first appeared in writing during the 13th century [29]. Table 2.3 below lists Tigrigna Fideles. The columns are assigned to the seven vowels of Tigrinya and the rows are assigned to the consonants. A Tigrinya syllable may consist of a consonant-vowel or a consonant-vowel-consonant sequence.

Pronunciations of the Tigrigna syllables are the same as that of Amharic with little modifications.

Table 2.3. Tigrigna Fidels [29]

	ä	U	i	a	E	(ə)	O
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
ḥ	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
s	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
r	ረ	ሩ	ሪ	ራ	ራ	ር	ሮ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
š	ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ
k'	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
kw'	ቈ		቉	ቊ	ቋ	ቌ	
x'	ቐ	ቑ	ቒ	ቓ	ቔ	ቕ	ቆ
xw'	ቈ		቉	ቊ	ቋ	ቌ	
b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
v	ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ
t	ተ	ቱ	ቲ	ታ	ቲ	ት	ቶ
č	ቸ	ቹ	ቺ	ቻ	ቼ	ች	ቾ
h	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
n	ነ	ኑ	ኒ	ና	ኔ	ነ	ኖ
ñ	ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ
'	አ	ሉ	ሊ	ኣ	ሄ	እ	አ
k	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
kw	ኸ		ኹ	ኺ	ኻ	ኼ	
x	ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ
xw	ኸ		ኹ	ኺ	ኻ	ኼ	
w	ወ	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ
'	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
z	ዘ	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ
ž	ዘ	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ
y	የ	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ
d	ደ	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ
ǧ	ጀ	ጁ	ጂ	ጃ	ጄ	ጅ	ጆ

g	ገ	ገጽ	ገጽ	ገጽ	ገጽ	ገጽ	ገጽ
gw	ገጽ		ገጽ	ገጽ	ገጽ	ገጽ	
t'	ጠ	ጠጽ	ጠጽ	ጠጽ	ጠጽ	ጠጽ	ጠጽ
č'	ጨ	ጨጽ	ጨጽ	ጨጽ	ጨጽ	ጨጽ	ጨጽ
p'	ጸ	ጸጽ	ጸጽ	ጸጽ	ጸጽ	ጸጽ	ጸጽ
s'	ጸ	ጸጽ	ጸጽ	ጸጽ	ጸጽ	ጸጽ	ጸጽ
s'	ፀ	ፀጽ	ፀጽ	ፀጽ	ፀጽ	ፀጽ	ፀጽ
f	ፈ	ፈጽ	ፈጽ	ፈጽ	ፈጽ	ፈጽ	ፈጽ
p	ፐ	ፐጽ	ፐጽ	ፐጽ	ፐጽ	ፐጽ	ፐጽ

2.3. Electronic Dictionary for Ethiopian Languages and the Challenge

A number of electronic dictionaries for several languages are becoming available over the WWW whether bilingual or multilingual. Huge companies working on linguistic area are also providing powerful dictionary applications with vast number of words of some languages for sale. Smaller yet powerful offline electronic dictionaries of some languages are also becoming available on small, handheld devices to enable the device users get instant access to the material.

Prior to the boom in desktop publishing in the late 1980s, electronic text in few Ethiopian languages existed as the product of experimental explorations into computer environs and was few in number [22]. As personal computers came down in price and word processing software became more practical and extensible the stage was set for local publishing to get underway in a significant manner. Currently, a number of documents and information written in local languages are available electronically over the WWW and some applications and software are being developed by local languages to be used on desktop computers. But the number of electronic dictionary applications that are developed for any of the Ethiopian languages is very much small in contrast to the other world languages.

Moreover, there exists no single dictionary application designed and developed for mobile devices that tried to use any of the Ethiopian languages even though these devices are widely used in the country. Several factors can be considered that might have contributed to this lack of local mobile applications. To mention some:

- The complexity of the scripts used to write texts. As the Ethiopic script constitute 34 core symbols each with six extensions, 44 labialized symbols, 20 numeral and 8 punctuation

symbols and also writing these characters involve combination of two or more English characters, it requires tedious work representing data written with this format to its electronic form.

- The resource consumption. The CPU cycle and memory consumption to store and retrieve texts written with Ethiopic script can also be very high compared to storing English or Latin data. As mobile devices are even more resource constrained like smaller storage capacity and CPU cycle, smaller screen size etc, developing applications of these languages bring even more challenging issue.
- The lack of appropriate mobile environment. As the mobile platforms used since recently are not designed to support the Ethiopic scripts, it was impossible for researchers and developers to think of designing an application with the local languages that uses this script.

These points and others have to be taken into consideration while thinking of designing local applications for mobile devices. Especially designing an electronic dictionary application, whether mono, bi, or multilingual on mobile devices is more challenging as the data to be stored needs to be large enough which consumes large memory size, the lookup for each word requires several CPU cycle and also the format of the displayed result needs to comply with the limited screen size of the device.

Chapter Three

Related Work

There are many online dictionaries; monolingual, bilingual, and even multilingual for different languages being available over the Internet for free or with some cost. Lots of researches had been carried out related to designing and implementing these dictionaries world wide and some of them will be discussed in this chapter.

3.1. Monolingual ED

Monolingual dictionaries use similar language to define a word. It defines a word in a different and elaborated form using the same language as the one used to write the word so that the user can clearly understand the different meanings and uses the word have in that same language. These kind of electronic dictionary are the easiest to construct as there is no need of translation to another language and hence require less memory space. Several works had been conducted regarding the design and implementation of monolingual electronic dictionary for different languages.

Pascale B., Jacques D., and Jean-Marie., [30] presented in their work the TLFi (Trésor de la langue française informatisé) monolingual electronic dictionary which is, as they claim, the most important computerized dictionary on French language. According to the authors, they have been able to insert in the text a very complex set of Extensible Markup Language (XML) tags so that every textual object is clearly identified and that the hierarchy containing these objects is clearly designed. The TLFi uses its own specialized software known as STELLA that, together with the XML structure, allows high level queries as well as hyper-navigation through and between different databases.

The design of a lexical resource focusing on German verb phrase idioms is presented in [31]. The paper tried to describe the properties of verb phrase idioms and created a resource that combines features of a dictionary, a grammar of idioms and a corpus. In this work an electronic knowledge base and a corresponding sub-corpus of examples for each entry was created via a structured annotations.

A very large Russian dictionary on the WWW is also described by [32] which is stated to be the first of its type in that it includes patronymic links for a given word. patronymic link is a link that

connects entries having similarity in letters. According to the authors, the dictionary contains more than 3 million links between 120,000 entries and the entries can be both single and multi-words. It operates with two main data structures: a list of entries and a set of links between them.

James McCracken in his work presented a project to develop a lexicon for use both as an electronic dictionary and as a database for a range of Natural Language Processing (NLP) tasks [33]. He proposes that a lexicon for such open-ended application may be derived from a human-user dictionary, retaining and enhancing the richness of its editorial content but abandoning its entry-list structure in favor of networks of relationships between discrete lexical objects, where each object represents a discrete lexeme-meaning unit. The dictionary chosen for his project was the Oxford Dictionary of English as the author believed it is a relatively high-level dictionary, intended for fluent English speakers rather than learners and also has some key features that make it particularly appropriate for enhancement as comprehensive electronic lexical database, usable both as a dictionary and as a resource for exploitation in NLP applications. The major steps followed in the work are:

- Decomposition of the dictionary as a flat list of entries, to be replaced by a set of lexical objects
- Population of each lexical object with a complete set of morphological and syntactic data and
- Classification connecting lexical objects to each other in semantic hierarchies and networks of domain relationships

In addition to these, there are also much more researches that focuses on the design of a monolingual electronic dictionary for several languages like Arabic, Chinese, Hebrew and the like on the WWW.

3.2. Multilingual ED

Bilingual dictionaries use another language to define a word. When the translation of the word is given in more than one language, the dictionary is termed as multilingual or Multilanguage dictionary. Multilingual electronic dictionaries often include a database of cross-referenced unilingual dictionaries with the use of pivotal language or other Interlingua techniques such as

ontology. Some of the research works that have been done on designing and implementing a Multilanguage ED of some natural languages have been reviewed on this project.

The ONOMASTICA project [34], which is a European-wide research initiative for the construction of a Multilanguage pronunciation lexicon of proper names, is one example of such works. It seeks to create a set of pronunciation lexicons of European names, including city and town names, street names, family names, company and product names in a machine assisted fashion. According to the report, preparation of the project lexicon was carried out by expert phoneticians with the help of some customized software. More than nine languages of the European Community are covered in the project including Danish, Dutch, English, French, German, Greek, Italian, Portuguese and Spanish. The goal of the project was to derive pronunciation dictionaries for up to 1,000,000 names per language in a semi-automatic way and to investigate the problems of exchanging national names amongst the partners to create a matrix of 'nativised' pronunciations for each foreign name in each other language.

English—Serbo-Croatian Electronic Dictionary [35] is another work which describes a bilingual and bidirectional on-line Serbo-Croatian (SC)-English dictionary that has been available on the Internet since 1999. According to the authors, the dictionary is a wide-coverage, up-to-date, bidirectional, and bilingual covering not only general, often used terms, but also over 8,000 computer and Internet terms, as well as healthcare and medical vocabulary, including useful abbreviations. The entries are grouped by semantic meaning and part of speech.

Svetlana Sheremetyeva [36] in his work focused mainly on presenting application tuned electronic dictionaries. The work presents the TransDict project that target on developing a multilingual electronic dictionary.

An Intelligent Multi-Dictionary Environment is an open, extendible multi-dictionary system which supports a translator in accessing adequate entries of various bi- and monolingual dictionaries and translation examples from parallel corpora and is presented on [37]. The author propose the MoBiDic electronic dictionary which is a multi-dictionary environment based on a client-server architecture consisting linguistic server, dictionary server and the client with the graphical user interface. The work uses a lemmatizer that provides the dictionary look-up module with the stem of the input word so that the user is not forced to only type the head root of a word. According to the author the work helps the user to also find all the multi-word expressions

containing the actual words' stem which would be impossible to find in traditional paper-based dictionaries.

Gábor Proszeky and Balázs KIS [38] also presented in their work a 'Context-Sensitive Electronic Dictionaries' that provides translations for any piece of text displayed on a computer screen, without requiring user interaction. To achieve their goal, they propose three phases in the work which are: text acquisition from the screen, morpho-syntactic analysis of the context of the selected word, and the dictionary lookup.

The Papillon project [39] is another research work that tried to establish a multilingual dictionary system on the Web. Its goal is to build a French-English-Japanese multilingual lexical database and to extract from it digital bilingual dictionaries. The project uses monolingual dictionaries of each language to build the multilingual or pivot dictionary by linking each sense or meaning of each entry of the monolingual dictionaries to one or more definitions (termed as acceptions) of the pivot dictionary. By doing so, the authors claimed that several monolingual as well as multilingual dictionaries can be extracted which can be used either by human or the machine.

There are also much more bilingual as well as multilingual electronic dictionaries online on the WWW developed for several languages spoken all over the world that can be used for free.

Despite all these works being done trying to establish an ED of several natural languages, there does not exist, as of our knowledge, a single research work that had tried to design and implement EDs for any of the Ethiopian languages. However there are few websites that host dictionaries of Amharic-English, Tigrigna-English and Afaan Oromo-English languages but are not in a much compiled way [44, 45].

3.3. Mobile ED's

In addition to the aforementioned research works that tried to build an electronic dictionary over the WWW for several world languages, there are also other works that had been done towards adopting these dictionaries on mobile devices like pocket PCs, PDAs and smart phones by different software companies.

The SlovoEd Multi-language dictionary for Nokia Communicator is such kind of software which has been developed for Nokia devices that consists several European languages [40]. It is stated as advantageous in that it encompasses great number of word entries, has low memory

consumption, the possibility to install several language pairs and convenient interface and color formatting of word entries. On top of these, the SlovoEd dictionary has also a sound module that will help the user to listen to the words' pronunciation.

ABBYY Lingvo x3 Mobile is comprehensive multilingual dictionary software that can be used by smart phones, PDAs and communicators released by ABBYY, one of the leading providers of document recognition, data capture and linguistic software [41]. According to the company, their product supports translation of words and phrases from English, German or Russian into German, English, French, Italian, Russian, Spanish and back. This mobile dictionary also includes additional useful features such as audio pronunciations and a tutor tool for vocabulary training that enables people to take advantage of the available time to improve their language and communication skills.

There are even much more EDs designed for mobile devices both as platform dependent and independent form that can be purchased from the publishing company or even can be used for free. Unfortunately, similar to the multilingual online dictionary, there is no dictionary software designed constituting any of the Ethiopian languages for mobile devices even though these mobile devices are being used widely by Ethiopians all over the world.

Chapter Four

Requirement Analysis

In the previous chapter several works regarding design and implementation of an ED for several natural languages has been discussed. Despite the large number of work available, few attempts had been done to build online ED for some of the Ethiopian languages. To mention some, the online Amharic-English dictionary found on [42] and the Tigrigna-English glossaries at [43] are available with some common words of each language. To our surprise, there does not exist, as far as observed, any single work that tried to present a multilingual dictionary of Amharic, Afaan Oromo and Tigrigna languages neither in paper format nor electronically, despite the existence of large number of these language speakers.

No single dictionary application for any of these three Ethiopian languages which had been designed to feet on mobile devices can be found on the market either. This is the first attempt to design and implement a four-in-one and multidirectional ED containing Amharic, English, Oromiffa and Tigrigna languages for any handheld devices and in particular for Android-based smart phones. As a result no previous work can be referenced and discussed here as there does not exist any.

In this chapter the functional and non-functional requirements of the proposed system will be described and modeled using the Object Oriented UML models.

4.1. Proposed system

4.1.1. Overview

The Amharic, English, Oromiffa and Tigrigna multilingual mobile dictionary is going to be used as a pocket dictionary that users can carry it everywhere and use it at any time easily. The idea of providing this dictionary application on the smart phones or any handhelds is so advantageous in that it takes only a few minutes to learn, even for the most techno-phobic, it's not price-prohibitive, the user need not have to invest in extra devices, nor waste time and energy which are all the limitations of paper dictionary from being used. The psychological stress of not remembering 'The Word' that we need is lowered too as we can easily search for the word on our

mobile phone and it can be an effective language learning tool and interesting application that can capture users' interest. In general, a user having this electronic dictionary on his/her mobile device gets an immediate feeling of control over his/her environment, and the other party in the conversation has immediate comprehension; hence providing general harmony and well-being. It is also a good teaching material for students and those who want to know the major three local languages spoken in Ethiopia together with the most widely used International language in the country.

The proposed system comprises four dictionary subtypes each of which having three branches; the Amharic dictionary subtype, the English dictionary subtype, the Afaan Oromo dictionary subtype and the Tigrigna dictionary subtype. Each subtype accepts input word from the user using the language the subtype is representing, which is the source language, and looks for its corresponding output or translation using the other three languages. The system allows the user to choose the type of dictionary to use from the list of the subtypes. It should also let the user type the word that he/she is looking for by providing the appropriate input method so that the user can type in the word with his/her own language. Moreover, the system is expected to display a list of suggestions that are hoped to match the word the user is going to look for when the user starts writing in order for the user to choose from the provided list the word he/she wants to look-up instead of writing the whole word. The proposed system should then search for the definition as well as pronunciation of the word and display the appropriate results in the chosen language and also allow the user to switch from the chosen language or dictionary subtype to another one. The overall work flow diagram of the proposed system is shown below in Figure 4.1.

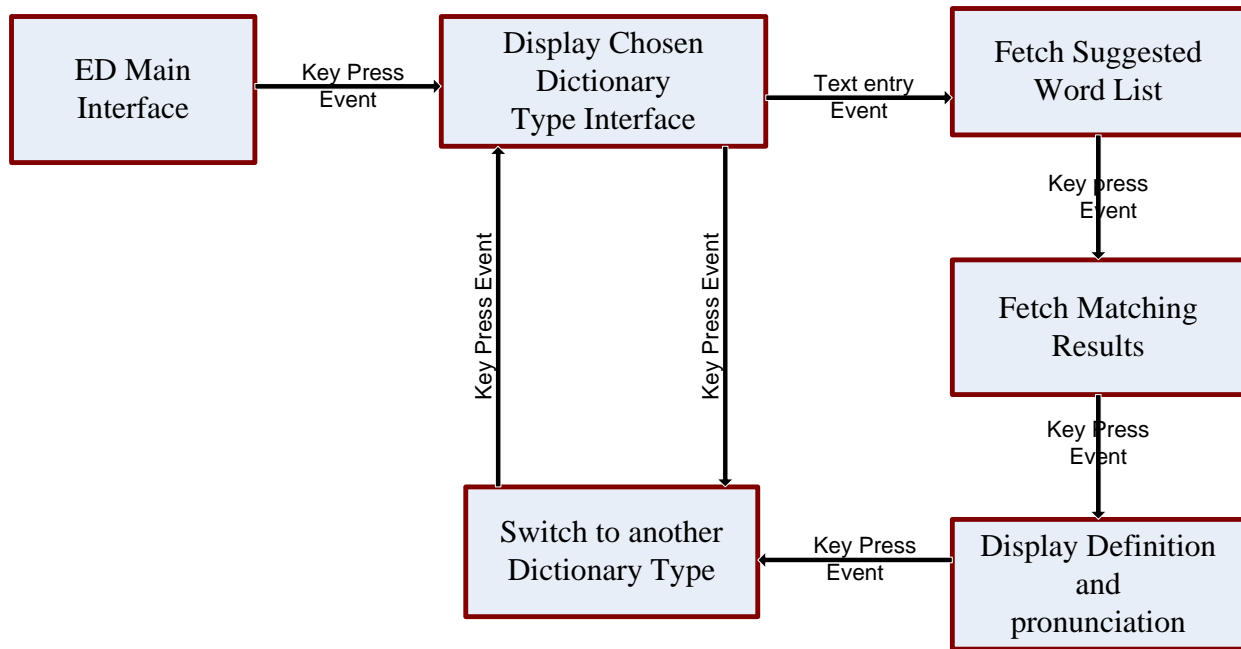


Figure 4-1 Work flow diagram of the mobile dictionary system

As Figure 4.1 shows, the system first displays the main dictionary interface on the screen listing the available dictionary subtypes as an option to choose from. As the user chooses his/her preference, the system will display the chosen dictionary type interface together with a soft keypad that can be used to write texts on the interface. The keypad can be either the Latin keypad or the Ethiopic keypad from which the user can choose the appropriate one depending on the dictionary type launched. Each dictionary subtype interface has a text entry field so that the user can type in the word to be translated using the language being familiar with. After the user starts writing, the system will automatically display list of words that are hoped to match with the word being typed to save the user from remembering the exact characters from which the word is constructed. After writing the word or selecting from the given list, the user presses one of the three search buttons to look-up for the word's definition and pronunciation in one of the three translating languages. Finally the system will search a match for the given query in the specified language and displays the definition of the word together with the pronunciation accordingly. The interface has also an option for the user that will let him/her to exit the current dictionary subtype and switch to another one.

4.1.2. Functional Requirements

The proposed system should provide the following functional requirements

- The system should allow the user to choose the preferred dictionary subtype
- The system should display both the Ethiopic and Latin soft keypad
- The system should let the user enter a query/word using the given keypad
- The system should be able to display list of suggested words that might match the word the user is looking for while the user is typing
- The system should search for the match of a given word and display its definition and pronunciation using the other chosen language
- The system should let the user to switch from the current dictionary choice to another one

4.1.3. Non-Functional Requirements

In addition to the functional requirements expected from the proposed system, there are also some non-functional requirements like:

- The delay time of the system to look-up for a word and also to switch from one subtype to another should be minimal
- The system has to be reliable in that it displays the exact match for a given query in a given language at anytime
- The system has to be more user friendly, easy to learn and use
- It also needs to comply with the limited resources of the mobile device

4.1.4. System Model

In this section the system's functionality is described using a use case diagram. Moreover, the dynamic and object model of the system is shown using sequence diagram and class diagram respectively. Finally, the different activities involved in the system are depicted using the activity diagram.

- **Use Case Model**

A use case can be defined as a way in which a user interacts with a given system in order to achieve some goal [44]. Use cases provide a means to capture system requirements, communicate with the end users and domain experts, and test the system. A use case describes a

function provided by the system that yields a visible result for an actor and an actor describes any entity that interacts with the system. This identification of actors and use cases results in the definition of the boundary of the system, that is, in differentiating the tasks accomplished by the system and the tasks accomplished by its environment.

Determining Use Cases

The following Use Cases has been identified from the system specifications and the use case diagram is shown in figure4.2 below.

- Choose Dictionary Type
- Get Suggested WordList
- Match Look-up
- Switch Type

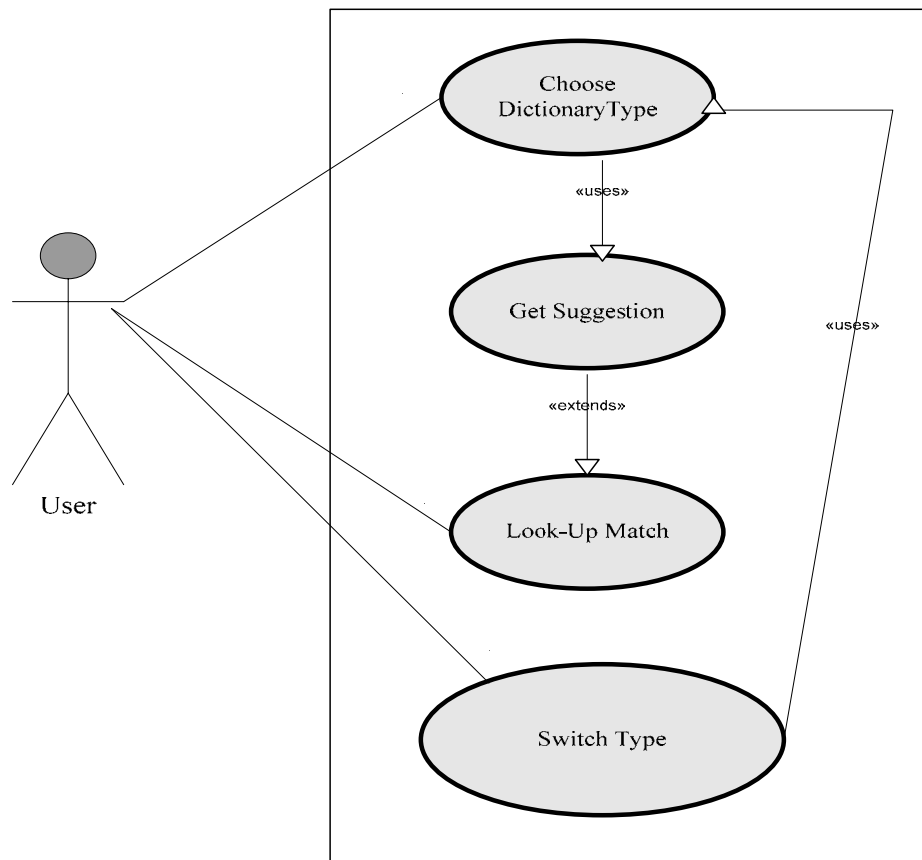


Figure 4-2 Use Case diagram of the proposed system

Use Case Description

Each of the identified use cases are briefly described below.

Table 4.1 Choose Dictionary Type use case description

Use Case Name	Choose Dictionary Type
Actor	User
Description	A use case that lets the user to choose the dictionary subtype to use
Pre-condition	The user must start or initiate the system
Flow of Events	<ol style="list-style-type: none">1. The user selects the type of dictionary needed to be used from the list of dictionary subtypes on the system2. The system changes the layout being displayed and loads the chosen subtype3. The system loads the keypad type chosen for text entry
Post-condition	Appropriate dictionary type interface is displayed

Table 4.2 Get Suggestion Use Case Description

Use Case Name	Get Suggestion
Actor	User
Description	A use case that displays list of suggested words for the user that might match his/her query
Pre-condition	<ol style="list-style-type: none">1.The user must select the type of entry dictionary to use2. The user must start typing a character so that there exist at least one character in the input box
Flow of Events	<ol style="list-style-type: none">1.The user starts typing a word2.The system stores the first typed character in array of characters3.The system searches for words that start with the same letter as the one inserted in to the array and creates a list holding the found words from the database4.It will continue searching the words database by adding the typed characters in to the array and refining its list

Post-condition	<ul style="list-style-type: none"> – The system displays list of words that are hoped to match user’s word in mind and the user chooses one, if needed – If no match is found, the system displays an error message to the user
----------------	---

Table 4.3 Switch Type Use Case Description

Use Case Name	Switch Type
Actor	User
Description	A use case that enables the user to switch from current dictionary type choice to another
Pre-condition	<ol style="list-style-type: none"> 1. The user must be in one of the dictionary subtypes found 2. The user must choose the dictionary type to switch to through a button press/menu
Flow of Events	<ol style="list-style-type: none"> 1. The user selects the desired dictionary type by a button press/menu 2. The system stops whatever its doing within the current environment 3. The system looks for the chosen type and fetches the layout 4. The user selects the keypad type to use
Post-condition	<ol style="list-style-type: none"> 3. The chosen dictionary subtype is loaded and displayed by the system

Table 4.4 Look-Up Match Use Case description

Use Case Name	Look-Up Match
Actor	User
Description	A use case that searches for a definition and pronunciation of a given word and displays the result, if found
Precondition	<ol style="list-style-type: none"> 1. The user must choose the dictionary type to use 2. The user must provide the word whose definition and pronunciation has to be looked for 3. The user needs to press a button after providing the input
Flow of Events	<ol style="list-style-type: none"> 1. The user types the word in the input box or picks a word from the suggestion list 2. The user then presses one of the four buttons that represent the translation languages 3. The system searches for a match of the given word in the given source language 4. The system look for the definition and pronunciation of the word in the target language
Post condition	<ol style="list-style-type: none"> 1. The system displays the definition and pronunciation for the given word if it is found in the chosen dictionary type 2. Otherwise, the system displays an error message notifying the user the word could not be found in the database

- **Sequence Diagram**

A sequence diagram of a system is used to formalize the behavior of the system and to visualize the communication among objects [44]. It is also useful for identifying additional missing objects that participate in the use cases. The diagrams below shows the sequence diagrams for the determined use cases of the system.

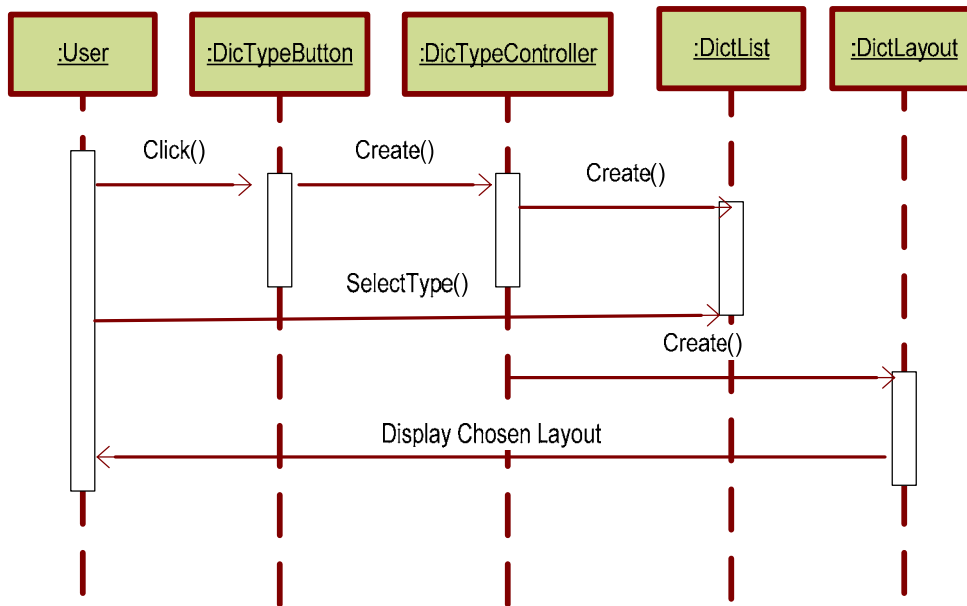


Figure 4-3 Sequence diagram for Choose Dictionary Type use case

Figure 4.3 shows the instance of choosing a Dictionary Type to use. The user activates the use case Choose Dictionary Type by interacting with the boundary object DicType button. The system then displays list of available dictionary types to choose from. The user selects one and submits to the system. Finally the system changes the layout and displays the appropriate interface of the chosen dictionary subtype.

Figure 4.4 below shows the sequence diagram of the second use case, Get Suggestion use case. This use case is initiated by the system automatically when the user starts typing a text into a textbox. The system then takes the first character typed and tries to find a word from the database that starts with the same letter as the one typed. It then creates an array of the characters typed in the textbox and continues to refine its search according to the characters typed and passed to the array. Finally, list of these suggested words is created and displayed on the interface so that the user can choose from the list the word he/she is interested in and perform the matching activity.

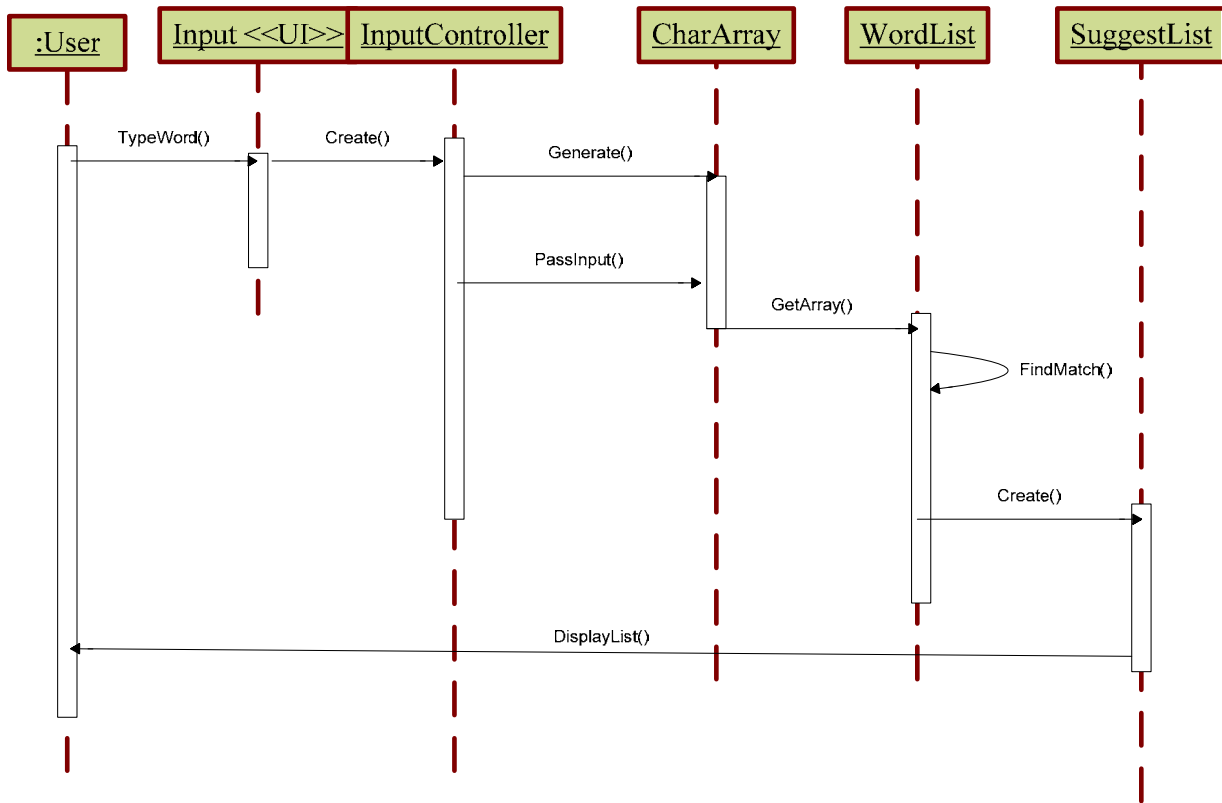


Figure 4-4 Sequence diagram for Get Suggestion use case

Figure 4.5, below, also describes the sequence diagram for the Switch Type use case. The use case is activated through a button click when the user wants to change the current dictionary type and switch to another. After the button is clicked, a list of subtype choices will be displayed and the user picks the desired one. The system finally updates the interface displayed to the chosen one.

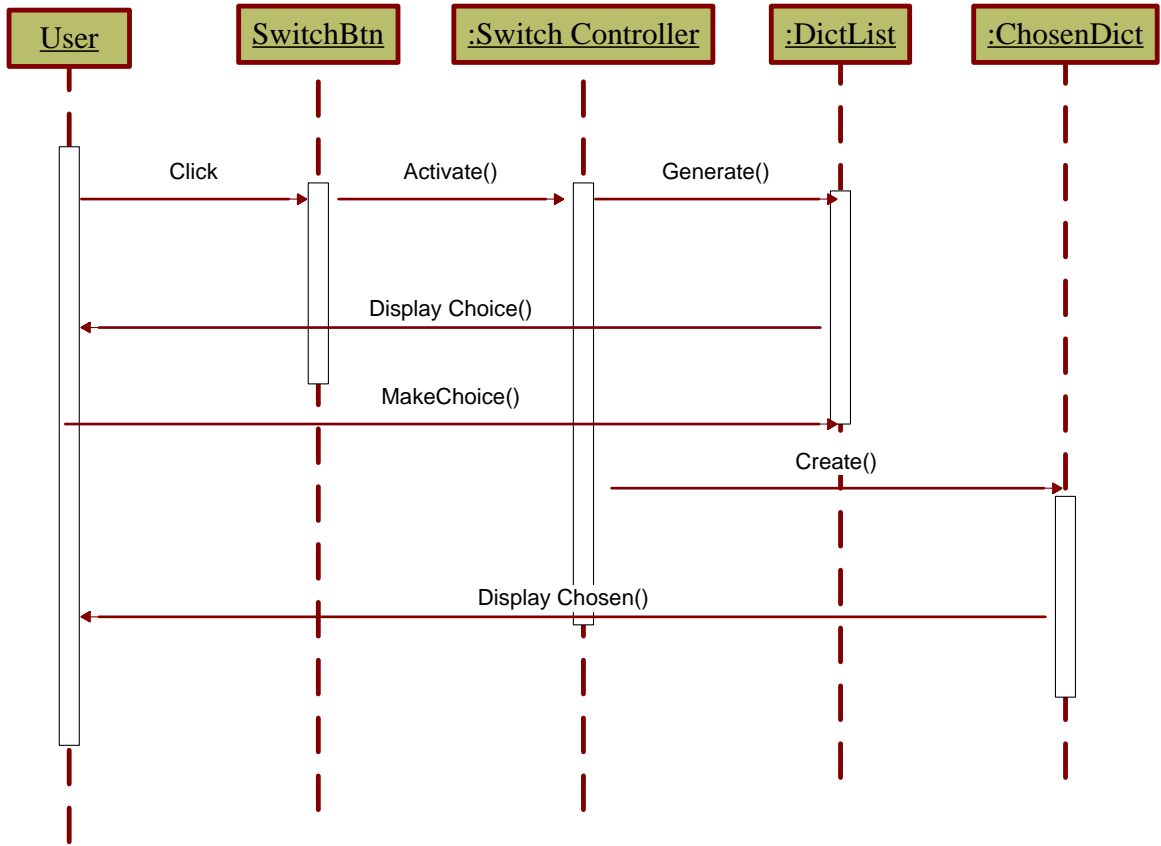


Figure 4-5 Sequence diagram of Switch Type use case

The last sequence diagram is drawn for the Lookup Match use case. This use case, as described earlier, is the one responsible for getting and displaying the definition and pronunciation of a given query. As can be seen from Figure 4.6 below, this use case is instantiated by the user after providing an input and pressing a button. The system will then look for the given input into the database and returns back the desired output related to the query for the user.

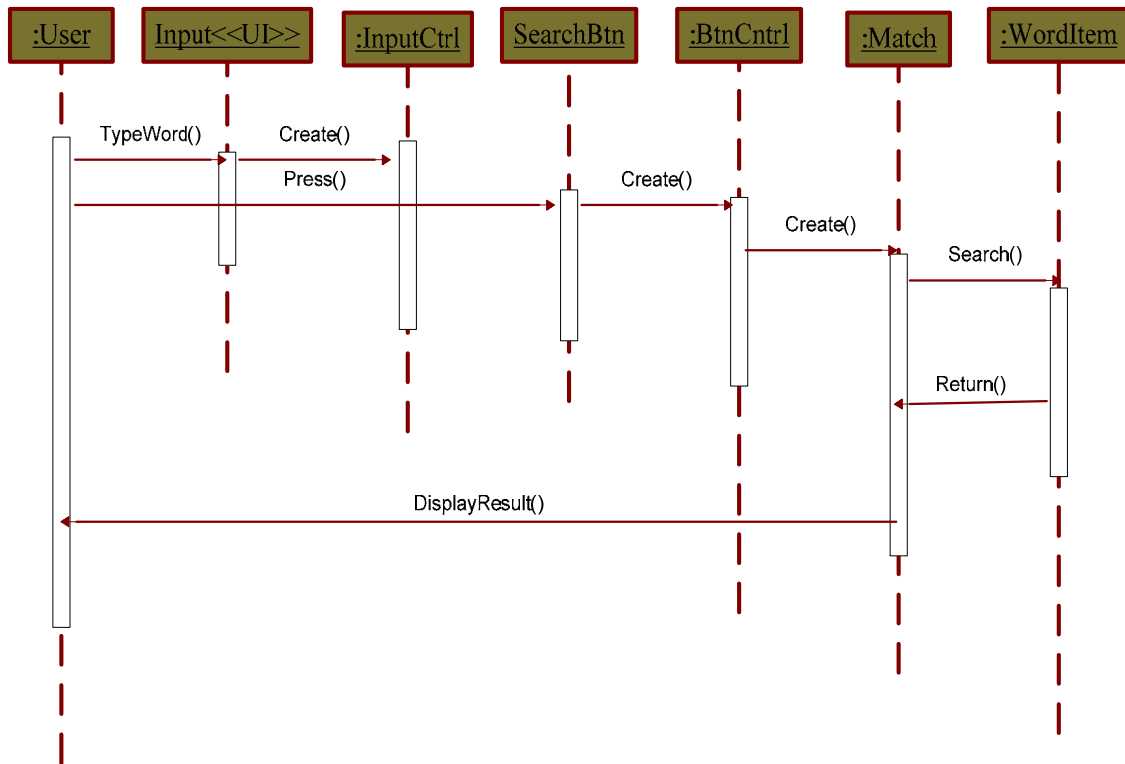


Figure 4-6 Sequence diagram for Lookup Match use case

- **Activity Diagram**

Activity diagram of a given system depicts the different activities that should take place in the system for its successful lifecycle and their concurrency. Activity diagrams are similar to flowchart diagrams in that they can be used to represent control flow and data flow. The activity diagram for the Ethiopic Multilanguage ED system is shown below in Figure 4.7 explicitly describing the flow of operations within the system. The activity of the system starts by acquiring choice input from the user on the startup interface which is used to load the entry dictionary subtype. The next stage will be displaying the chosen dictionary subtype and keypad type followed by word entry stage. The startup character of the given word will be then cross matched against the permanently stored data to display list of suggestion words which are similar to the given query. The user can either select one of the words in the suggested list or type with the keypad the whole word and press one of the four buttons to look for translation into another language. The system finally displays the translation and pronunciation of the selected word on the place provided using the selected language.

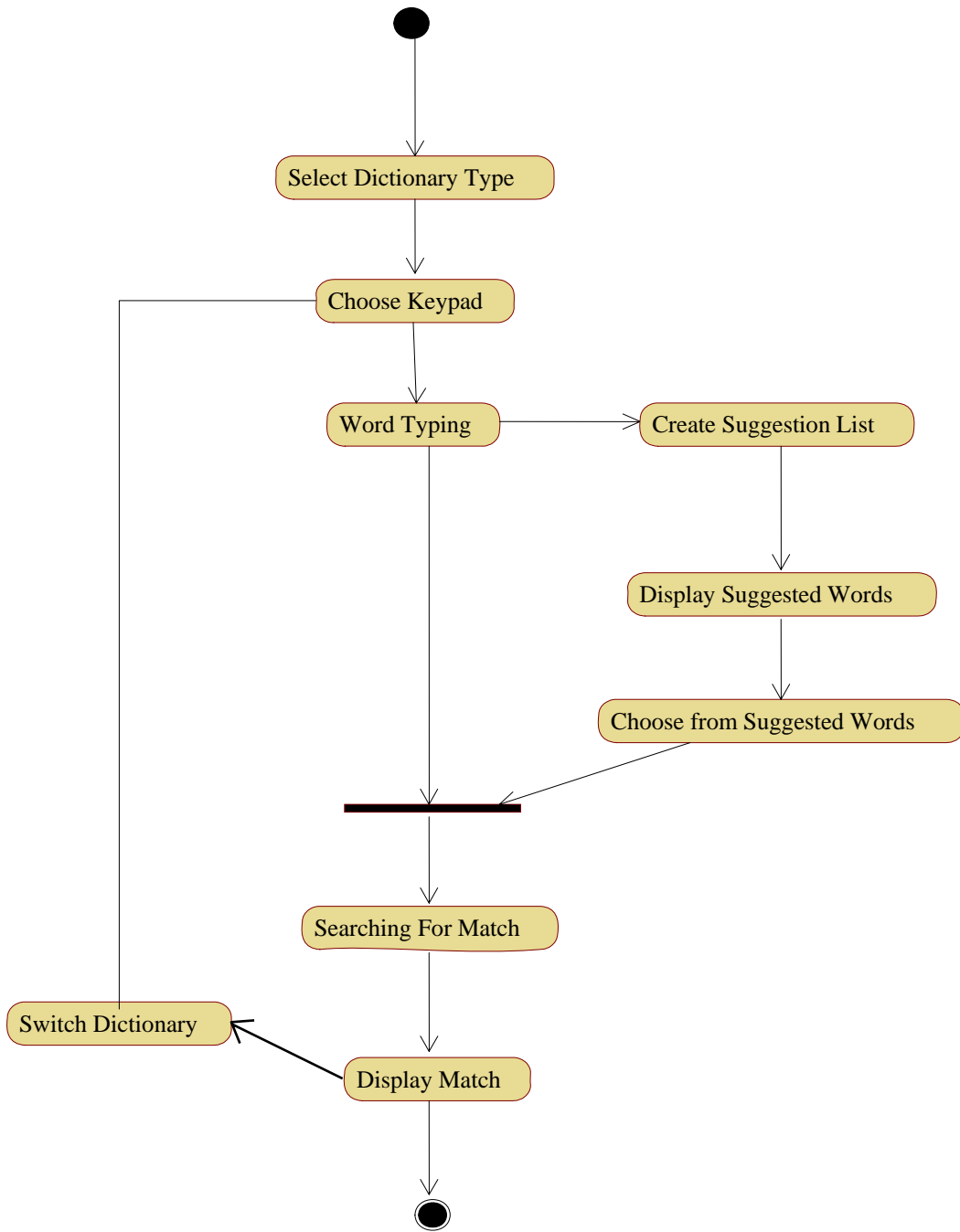
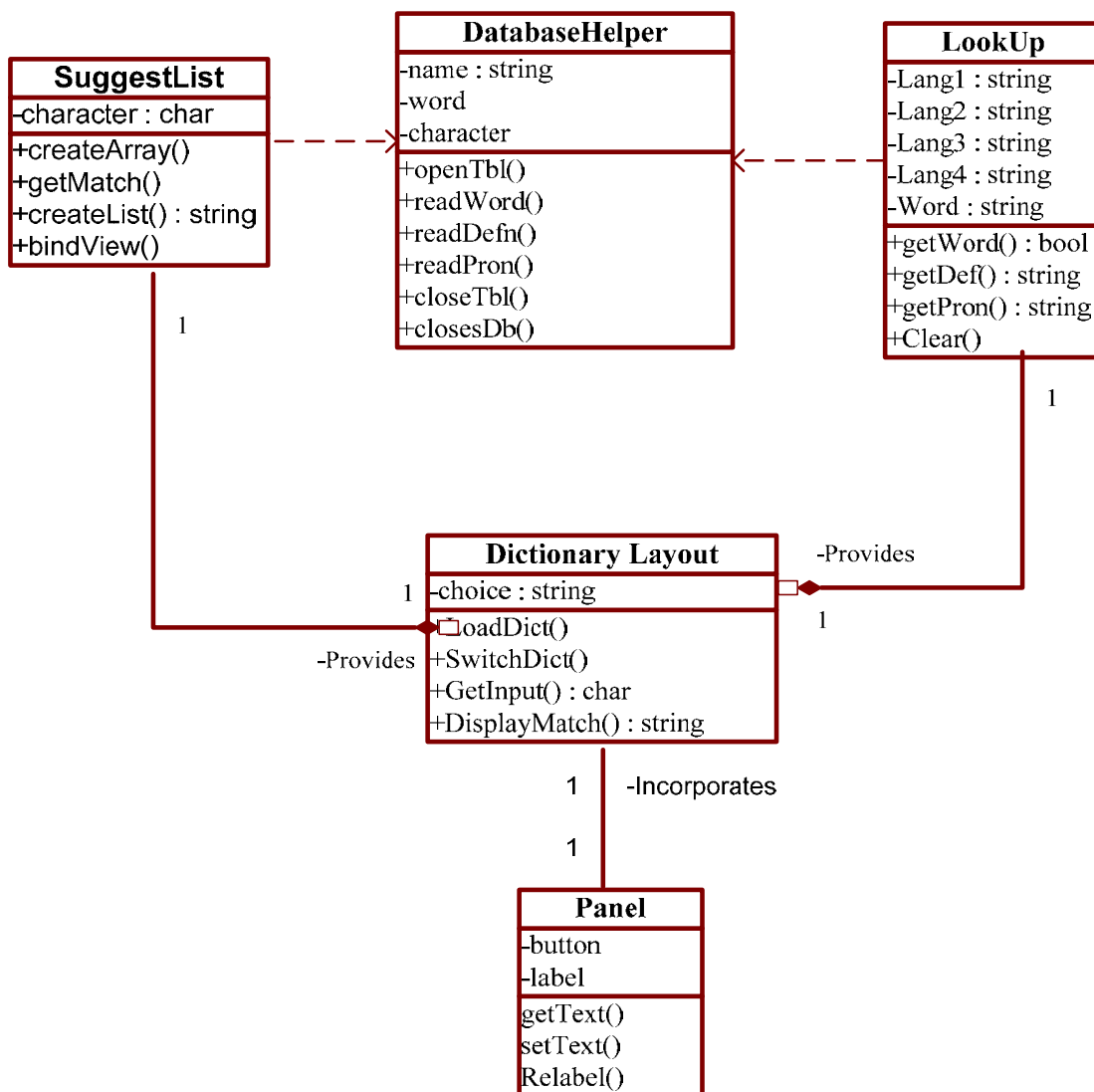


Figure 4-7 Activity diagram of the system

- **Class Diagram**

Class diagrams describe a given system in terms of objects, classes, attributes, operations, and their association. Classes are the primary building blocks of any object-oriented system. They represent the description of a collection of objects with common attributes, operations, relationships and semantics [44]. Well-structured classes are helpful in setting a properly balanced distribution of responsibilities within the system. Figure4.8 depicts the class diagram for the set of classes that are identified in our proposed system.



4-8 Class diagram of the system

Chapter Five

System Design

System design is the startup stage to get into the solution domain of a given system in a software development. After the requirement of a system is gathered and analyzed, then follows the design of the system. The design is all about stating the design goals of the system and subdividing the system into smaller parts so as to tackle the problem in a modular approach by taking into account the nonfunctional requirements and constraints described in the problem statement and requirement analysis sections.

In the previous chapter the functional as well as the non-functional requirements of the proposed system had been pointed out and analysis model had been constructed. In this chapter, the design model of the new system will be presented by specifying the design goals of the system, presenting the system's architecture and decomposing the whole system into sub systems.

5.1. Design Goals

The design goals of the system are derived from the nonfunctional requirements of the system which are pointed out in the previous chapter. They represent the desired qualities that the system should have by providing a consistent set of criteria that should be taken into account when making design decisions.

When designing an application for mobile devices, like the smart phones, one should be very careful and thoroughly consider the resource constraints on the devices like their data storage capacity, request processing speed, limited battery life and their limited screen size. There are always some contradicting requirements like space and speed while developing software for such devices and it is very important to balance these conflicting requirements.

The design goals of the proposed system take into consideration the following criteria.

5.1.1. Performance criteria

The performance criterion of a given system mainly considers the speed, throughput and memory requirement of the system.

Response time: the proposed system has to display the list of suggestion words just after the user typed the first letter of the query. Moreover, it has to search for a given word in the given language and provide the definition and pronunciation in the required language without noticeable delay. Switching from one dictionary subtype to another should also be done with minimal delay.

Throughput: the system must display the initial suggested list of words as the user types the first letter of the word before the next letter is typed and continue refining the list as the user continues to provide input characters.

Storage area requirement: any dictionary application requires sufficient storage space in order to be rich in resources and serve as a true reference material of a language. When it comes to a multilingual and multidirectional dictionary, this space requirement becomes even worse as each word of each language comes with its definition as well as pronunciation in all the languages considered. However, the currently available mobile devices are fabricated with a storage memory space capable of storing large amount of data. As a result, data storage space requirement for the proposed multilingual ED system will not be a huge constraint as it is going to be deployed on smart phones with enough memory size to store the necessary data of the application.

5.1.2. End User Criteria

End user criteria specifies the qualities of the system from users point of view

Usability: Usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

The question of usability is considered as one of a determining factor for users' acceptance of a given system. The system shall be designed and developed in such a way that it is easily understood and used by any mobile phone user. Its layout needs to match the layout the user is familiar with. Moreover, it shall also be more users friendly by providing guidelines and appropriate messages when an error occurs.

5.1.3. Maintenance Criteria

The maintenance criteria determine and discuss the difficulties that the system might face while being in use. The question of adding components to the system, modifiability and portability are the ones to be considered in the maintenance criteria.

Extensibility: the system shall be expandable and able to incorporate dictionary of other languages as well whenever needed without affecting the existing ones. Its functionality and performance need not be also affected due to the additional components.

Portability: the system shall work in any mobile device with no android platform as well as desktop computer with Java support with little or no modification as there might be a need to deploy this application on desktop computers, smart phones and other mobile devices with different platform and Operating System.

Modifiability: the system needs to be easily modifiable. It shall enable for any change in its design and configuration with minimum effort whenever the need arises.

5.1.4. Dependability Criteria

Dependability criteria of a system deal with the system crashes, their consequences and the effort needed to minimize system crashes. Security risk, safety issues and the availability of the system to the user are the main points being considered in these criteria.

Reliability: the proposed system needs to be consistent by performing the required functionality in any desired moment and providing the desired output correctly and consistently.

Availability: as the system is designed to work on mobile devices, it should always be active and functioning until the battery life ends.

5.2. Architecture of the Proposed System

Separation of layers is very crucial for any application structure. The architecture of the system determines the type of interaction among the components/layers of the system. The general architecture of the proposed system is broken down into three major layers as shown in Figure 5.1 below. These are the View Layer, the Controller Layer and the Persistence Layer each of which are briefly described as follows.

The View component is the one held responsible for providing explanation and description about the developed system to the target environment, and for the mapping of graphics onto the device. It is capable of rendering the contents of the persistence layer to the display surface by managing the graphical and textual output portion of the display allocated to the application. XML specifications about the view of the different available user interfaces are stored in the main.xml and dictionarytype.xml components of the layer defining the attributes included in each dictionary subtype activity.

The controller is the means by which the user interacts with the application. It acquires input from the user and handles the way the view and the persistence layer take action on the given input. It is responsible for mapping end-user action to application response by interpreting the mouse and key press event from the user, and commanding the bottom layer and/or the upper layer to change as appropriate. The user interacts with the system through several events like a button click, menu choice and a text field entry. Depending on the user interaction and the provided output from the persistence layer, the controller responds by selecting the appropriate view from the view action.

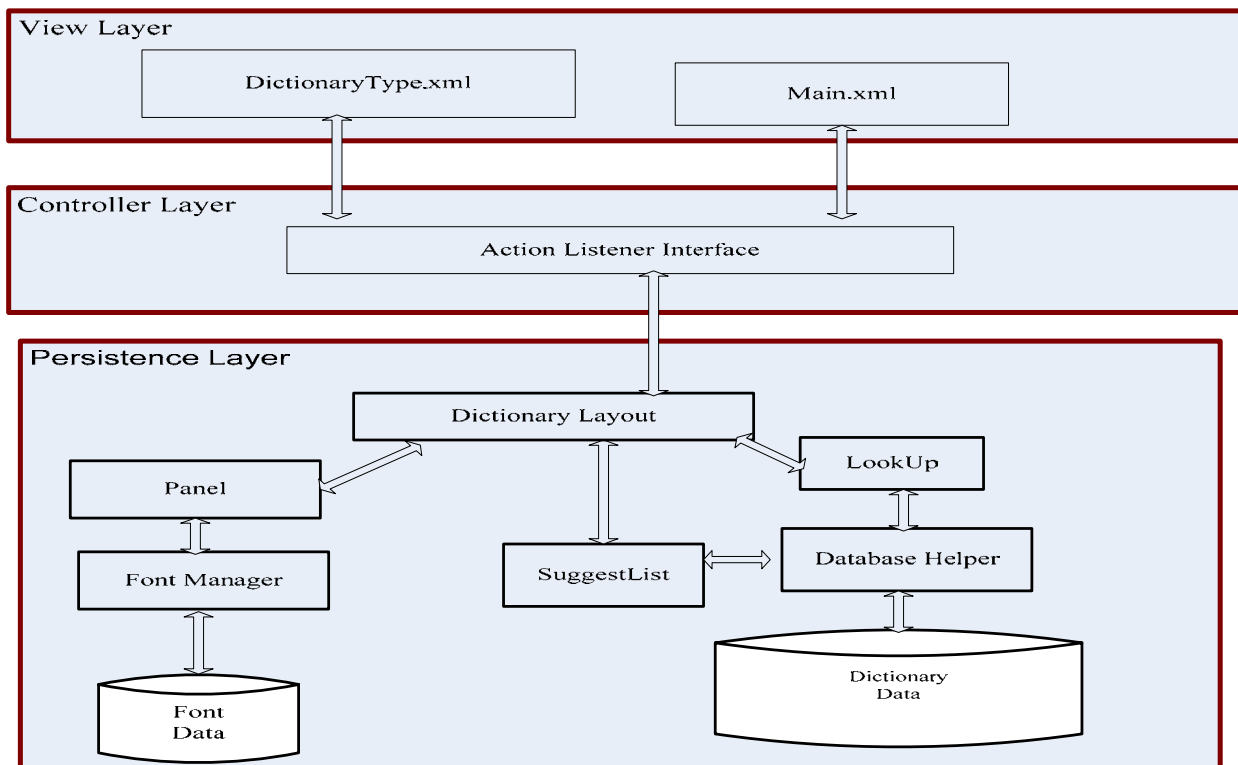
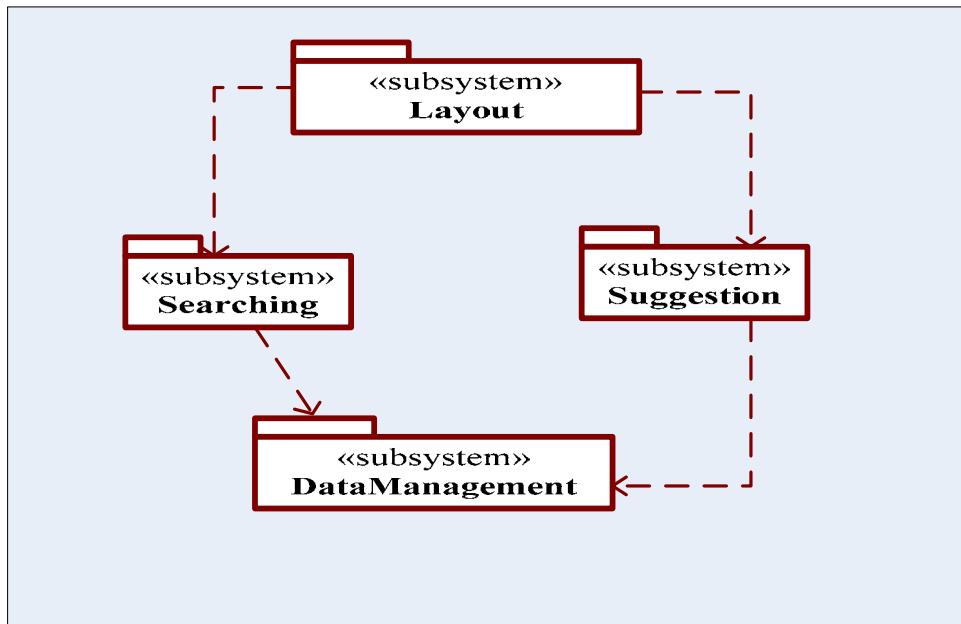


Figure 5-1 System architecture of the Ethiopic multilingual dictionary system

The persistence layer is the layer that manages the state of the application and conducts all the required functionalities. It holds the behavior and the data of the application domain and responds to requests made for information. This layer accepts word entry from the UI on the view layer, generates the desired result performing the necessary actions and finally passes the obtained result to the UI. To hold the application's data and perform the expected task, the layer uses two static data storages. The font data, which is used to load the different font types supported by the system, is stored on an external storage media of the device while the dictionary data that holds all the words and their translation languages is stored within the application's database system. This is the layer where all the classes defined in the project reside in.

5.2.1. Subsystem Decomposition

The purpose of subsystem decomposition is to breakdown a given system into smaller systems, which are loosely coupled with each other while each broken-down system has strong cohesion among its components. This reduces and avoids complexity in the system and makes easy tackling all the problems in the system. In order to simplify and minimize the complexity of the solution domain, the proposed Ethiopic multilingual ED system is decomposed into four subsystems namely the Layout subsystem, Searching subsystem, Suggestion subsystem and DatabManagement subsystem as shown in Figure 5.2 below.



5-2 Subsystem decomposition of the Ethiopic multilingual dictionary system

Layout Subsystem

This subsystem is the entry point of the system which is responsible for handling all the user interaction with the application. Its main task is to create and manage the different layouts of the application based on the available dictionary subtype. There are five different layouts, each for the four languages included in the system, to be managed by this subsystem. The subsystem loads the initial layout, which is the startup activity of the system and listens to the user action on that layout. This activity displays the available dictionary subtypes and lets the user choose from the given list. The other four activities represent the four languages included in the application and are loaded depending on the choice made by the user from the startup activity. Each of the four activities has a text entry field and buttons to be triggered by the user and provide an input to the system. It is the duty of this subsystem to get the user inputs and pass it to the searching as well as the suggestion subsystem and finally populate each layout with the acquired result from each subsystem. This subsystem packages the DictionaryLayout and the Panel class determined in the class diagram of the proposed system.

Suggestion subsystem

This subsystem is the one that creates and displays list of suggestion words from the database that are assumed to match the query in mind when the user starts providing an input. It accepts sequence of characters from the Layout subsystem and creates an array of the given characters. It will then call the DataManagement subsystem to retrieve all the root words in the database with similar head character as the first character in the array. The subsystem then calls the Layout subsystem and passes its output for the subsystem. It will continue consulting the DataManagement subsystem and refine its output by crosschecking every character that has been passed to it with the persistent data stored in the database. The SuggestList class is included under this subsystem

Searching subsystem

The searching subsystem is the third subsystem of the application and accepts an input from the Layout subsystem and consults the DataManagement subsystem for a match of the acquired input. The inputs provided for this subsystem from the Layout subsystem are texts which can be obtained either directly from the user or from the output of the Suggestion subsystem and a

button click event. It will then call the DataManagement subsystem and search for the match of the given input query by traversing at the first column of each data table to see if it exists. If the query is found to match one of the column values, its corresponding equivalent translation and pronunciation will be looked up. During the translation of the query, two approaches had been used by the subsystem. The first one is to select all equivalent translations of the query which will then pass the three languages translation and pronunciation of the query to the Layout subsystem. The second approach is to only select specific results based on the chosen translation language. This subsystem consists of the LookUp class.

Data Management Subsystem

This subsystem handles any database interaction and represents the persistence data of the system. It packages the DatabaseHelper class and is responsible for writing the whole dictionary data to the database and fetching them from the database as needed. The Searching and Suggestion subsystem both requests this subsystem only for reading operations to retrieve list of words, their respective definitions and pronunciations.

5.2.2. Persistent Data Management

The proposed system creates a database file and stores words for each dictionary subtype permanently for the four languages before the system is activated. To handle the database creation and management, the open source light weight SQLite DBMS is used in this project as it is already baked-in with the Android platform and is so space-efficient which makes it preferable to be used on devices like smart phones whose storage capacity as well as CPU cycle is limited in nature. SQLite is a very popular open source embedded database combining a clean SQL interface with a very small memory footprint and decent speed [45]. It uses a dialect of SQL for data definition, data manipulation and queries.

The database of this system contains four tables each for storing data of the four languages. Each table has seven attributes that holds set of records for root words of the corresponding language, translations for each root word using the other three languages and pronunciation of each translated word with the root word language. The Amharic dictionary table, for instance, holds Amharic root words, their meanings or senses in English, Afaan Oromo and Tigrigna and pronunciations for the English and Afaan Oromo definitions. Pronunciation for Tigrigna language is skipped in the Amharic dictionary table as both Amharic and Tigrigna languages are

written with the same Ethiopic script and it is assumed anyone reading Amharic words can also read Tigrigna words and also no resource was found on how to pronounce Amharic words in Tigrigna and vice versa. Moreover, an early decision had been made to avoid listing and storing derived forms of words for each language and only the root words are available. Each entry or root word in each dictionary table contains one or more senses or meanings in the other languages depending on the usage of the word in its language. For example the word “ገና” in Amharic has two meanings, ‘not yet’ and ‘Christmas’ in English, so this word is linked with two translations and pronunciations for each language.

The databaseHelper class is the one responsible for creating the database and the tables and also for populating each table. The records for all the tables are inserted both manually from published paper dictionary books as well as available softcopies of some dictionaries by first storing the whole data in a text file as Unicode text format and importing the content to the database tables. The Genius English-Amharic Dictionary by Hinsene Mekuria[46] and DUNGOO Oromo-Amharic-English dictionary again by Hinsene Mekuria [47] , which are both in hardcopy format and the Tigrigna-English, English-Tigrigna Dictionary[48] which had been compiled as a pdf document and made available online are used as a source of data on this project. As it was difficult to get a dictionary of Amharic-Tigrigna as well as Afaan Oromo-Tigrigna languages, translation of Amharic words into Tigrigna and Afaan Oromo words into Tigrigna are done taking their English meaning as a pivotal word, that is, taking the English translation of the Amharic as well as Afaan Oromo words and then look into their corresponding Tigrigna words in the English-Tigrigna dictionary. The overall database design with the four tables and their attributes is shown below.

English Table	
Word	text not null
_id	int primary key autoincrement
Adef	text not null
Apron	text not null
Odef	text not null
Opron	text not null
Tdef	text not null
Tpron	text not null

Amharic Table	
Word	text not null
_id	int primary key autoincrement
Edef	text not null
Epron	text not null
Odef	text not null
Orpron	text not null
Tdef	text not null

Tigrigna Table	
Word	text not null
_id	int primary key autoincrement
Adef	text not null
Edef	text not null
Epron	text not null
Odef	text not null
Opron	text not null

Afaan Oromo Table	
Word	text not null
_id	int primary key autoincrement
Adef	text not null
Apron	text not null
Edef	textnot null
Epron	text not null
Tdef	text not null
Tpron	text not null

Figure 5-3 Table designs of the four languages

5.3. Algorithm design

In this project, a simple algorithm has been developed to display list of suggested words that are expected to match the query the user is going to look for as shown below in Table 5.1.

Retrieval of suggested words from the Multilanguage electronic dictionary is done by matching the keyword the user provides against entry words within each dictionary. However, as can be seen from the table, instead of seeking an exact match of keyword against dictionary entry, the keyword is broken down into character sequences and the longest possible matches are obtained and displayed followed by the next longest and so on by refining the lookup procedure based on the character sequences of the keyword.

Thus, initially the first character of the keyword will be matched against those dictionary entry with the same initial character. Those obtained entries with the same initial character are displayed on the display view. But if there are other characters provided by the user, the system filters out the displayed entries by crosschecking the second character of the input with the

second character of these entries. This procedure will go on for all the character sequences found in the keyword and the obtained result will be filtered out accordingly.

Table 5.1 algorithm for displaying list of suggested words

```
//define autocomplete input area
//define getCursor, list, bindView and refine methods for accessing dictionary entry, storing
matches, displaying the matches and filtering out the displayed results
input:      - Source language
            - Character sequence
read source language from the input
//call getCursor method to open and read the database table of the source language
get input character from input field
count: number of records of the opened table
input: array of characters to hold input characters
record: array of records retrieved from the opened table
size: number of character in the array
  J ← 0 //the first record in the table
Do
{
  I ← 1
  //compare first character of the array with the first character of jth record of the table
  If(record[j][i]==input[i]){
  //add the obtained result to the result set
  List(record)
  }
  UNTIL(j>=count)
  //display obtained matches
  bindView(list)
  if(size>1)
  k ← 2 // index of characters after the first character
  Do
  {
```

```
//compare kth character in the array with the kth character of the results in the list and remove those that fails to qualify
If(list[k]!=input[k])
Filter(list)
BindView(list)
}
WHILE(k<= size)
```

output: list of suggested words to choose from and perform more search

As can be seen from the table above, the system accepts a single character from the user and goes through the given table to lookup words that start with the same letter as the one provided. If the user found the word he/she was going to write from the list displayed by the system and select one, then the lookup will be stopped. Otherwise if the user ignores what has been displayed and continues typing, the system refines its search by only looking into the displayed list and repopulates the list with more refined outputs. The user can then choose from the list the word to be translated or can provide the keyword by writing all the character sequences of the word.

This is a very crucial feature for any dictionary application as it avoids the stress on the user of remembering the character sequence from which the keyword is made up and makes life easy for the user by providing the word to be translated quickly.

Chapter Six

Implementation of the Multilingual Dictionary

This chapter provides the implementation details of the Ethiopic Multilingual Dictionary System whose design has been provided in Chapter Five. An implementation of the designed Ethiopic Multilingual ED is carried out for Smart phones with Android Platform. An insight to the development environment of the system will be presented in section 6.1 while the tools used in developing the system and the developed system itself will be briefly discussed in the following two sections.

6.1. Development Environment

Android is an open source software stack that includes an OS, middleware, and key applications along with a set of API libraries for writing mobile applications that can shape the look, feel, and function of mobile handsets. It was first released in 2007 and is owned by Google and OHA [5]. It is made up of several necessary and dependent parts including the following [49]:

- ❖ A hardware reference design that describes the capabilities required of a mobile device in order to support the software stack
- ❖ A Linux operating system kernel that provides the low-level interface with the hardware, memory management, and process control, all optimized for mobile devices
- ❖ Open source libraries for application development including SQLite, WebKit, OpenGL, and a media manager
- ❖ A run time used to execute and host Android applications, including the core libraries that provide Android specific functionality.
- ❖ An application framework that agnostically exposes system services to the application layer, including the window manager, content providers, location manager, telephony, and peer-to-peer services
- ❖ A user interface framework used to host and launch applications
- ❖ Preinstalled applications shipped as part of the stack and
- ❖ A software development kit (SDK) used to create applications, including the tools, plug-ins, and documentation

The Android Platform supports Internationalization by providing an engine that supports Unicode text layout and glyph rendering, which is very crucial for fonts different from Latin, like the Ethiopic fonts. It also provides the opportunity to create mobile phone interfaces and applications designed to look, feel, and function exactly as you image them. Moreover, all Android applications, whether native or third-party, have equal standing and are written using the same APIs and are executed on the same run time [45]. Users can remove and replace any native application with a third-party developer alternative. Hence, it is the appropriate mobile platform for developing and deploying any Ethiopian software on mobile devices and also the main reason to be chosen for this project.

6.2. Development Tools

The Android SDK includes several tools and utilities to assist in an application development for Android platform. Of all the development tools the Android Emulator and the plug-in tool which are used to integrate Android development with a chosen integrated development environment (IDE) are the major ones in the SDK and hence briefly described as follows.

Android Emulator:

The Android SDK comes packed with a mobile device emulator that runs on a computer or host machine and mimics all of the hardware and software features of a typical mobile device, except that it can not receive or place actual phone calls [49]. It enables us to develop and test our own Android application with out the need for the real device. It provides a variety of navigation and control keys, which can be triggered using mouse or keyboard to generate events for an application. It also provides a screen in which an application is displayed, together with any other Android applications running. The emulator supports Android Virtual Device (AVD) configurations that let us specify the Android platform that we want to run on the emulator and other options. The emulator also includes a variety of debug capabilities, such as a console from which one can log kernel output and simulate application. A typical Android Emulator with the Android 2.0.1 platform is shown in Figure 6.1 below.

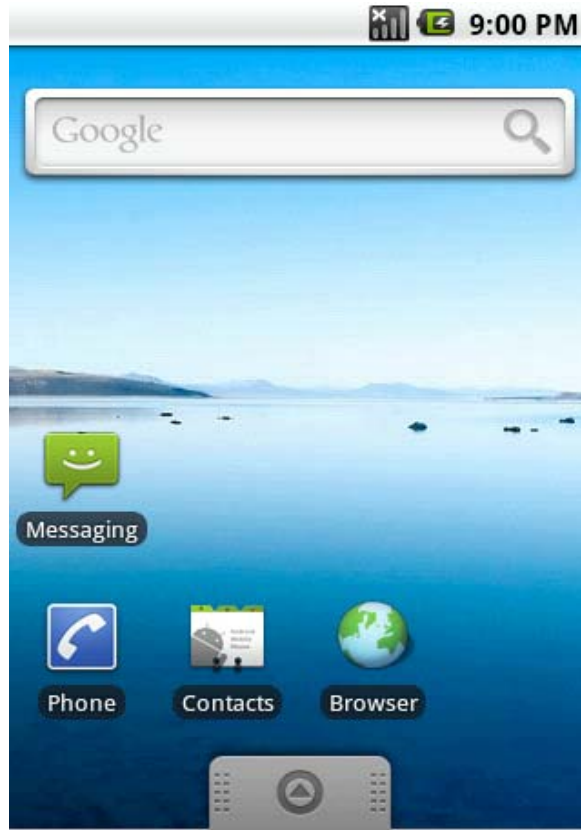


Figure 6-1 Android 2.0.1 platform Emulator

Eclipse IDE plug-in tool:

Android applications are written with Java programming language and the core Android libraries include most of the features from the core Java APIs. In order to develop and implement any Android application on a host machine, three tools need to be available on the system. These are a copy of the Android SDK, the Java development kit (JDK) and any Java IDE.

The Eclipse Java IDE, which is a free, easy to use and fully featured IDE, is used in this project as it is well supported by Android and the configuration is easier. Installation of an Eclipse ADT plug-in is also required as it simplifies the Android application development by integrating the developer tools, including the emulator directly into the IDE. It makes creating, testing, and debugging applications on the host environment faster and easier before uploading the application onto the real device. After successful creation of the android application on the Eclipse IDE, an Android Virtual Device (AVD) needs to be created for running the application on the emulator.

All the necessary tools described above are downloaded and installed on the host machine with the Windows OS for the successful implementation of the system.

Figure 6.2 below shows a screenshot of the Eclipse IDE with the ADT plug-in installed taken while the project is being developed.

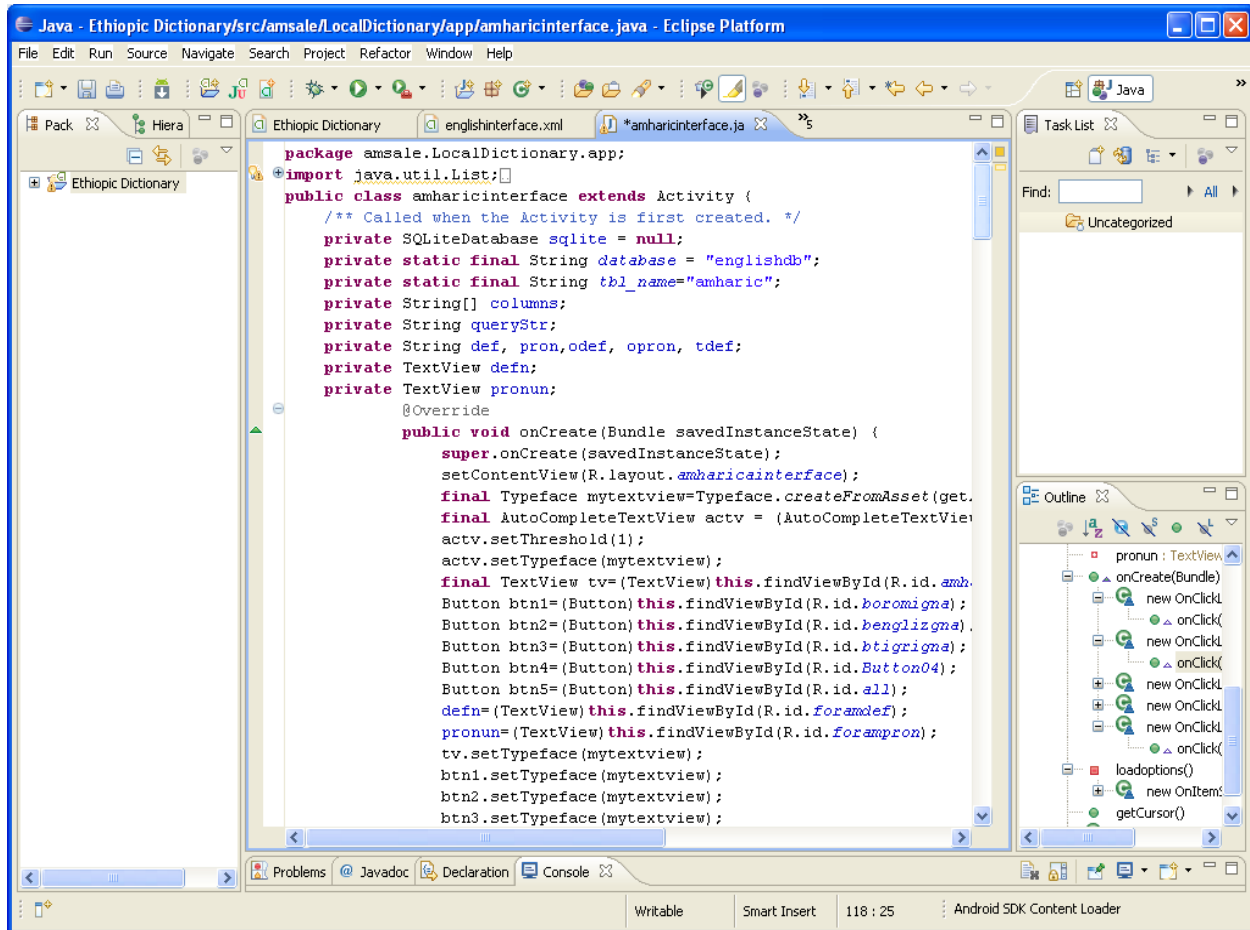


Figure 6-2 Eclipse IDE with the ADT plug-in-installed

6.3. The Ethiopic Multilingual Dictionary Application

The Android SDK provides the tools and APIs necessary to begin developing applications on the Android platform using the Java programming language. For this project, the implementation is developed using the Android SDK 2.0.1 platform and the Eclipse IDE with the ADT plug-in installed. After successfully creating and deploying the system onto the Android Emulator, the application will be available on the desktop of the emulator or the smart phone represented by an icon, just like the other native as well as third-party applications are presented, as shown in

Figure 6.3 below. In order for the user start using the application, he/she needs to click on the icon representing the application using a mouse if on the emulator or by either pressing the icon directly with hand from the screen of the Smart phone or using the navigation keys on the phone to traverse and pressing the enter key when reach to the application's icon.



Figure 6-3 The Ethiopic Multilanguage electronic dictionary icon

As stated earlier, every Android application is made from set of activities that interact with each other. The Amharic-English-Oromiffa-Tigrigna multilingual dictionary application also consist a total of five activities. Once the system is started, the first activity, which is the main activity of the system, will be activated. This activity displays a welcome note using the four languages and list of the dictionary subtypes to choose from and use. For the successful display of the Amharic texts, the Nyala.ttf font is being downloaded and used in this project which is then stored in the SD card of the smart phone.

The user then selects one of the listed options on the startup activity to launch his/her preferred dictionary subtype, which is the second activity. The screenshot of the startup activity is shown below in Figure 6.4 and Figure 6.5.

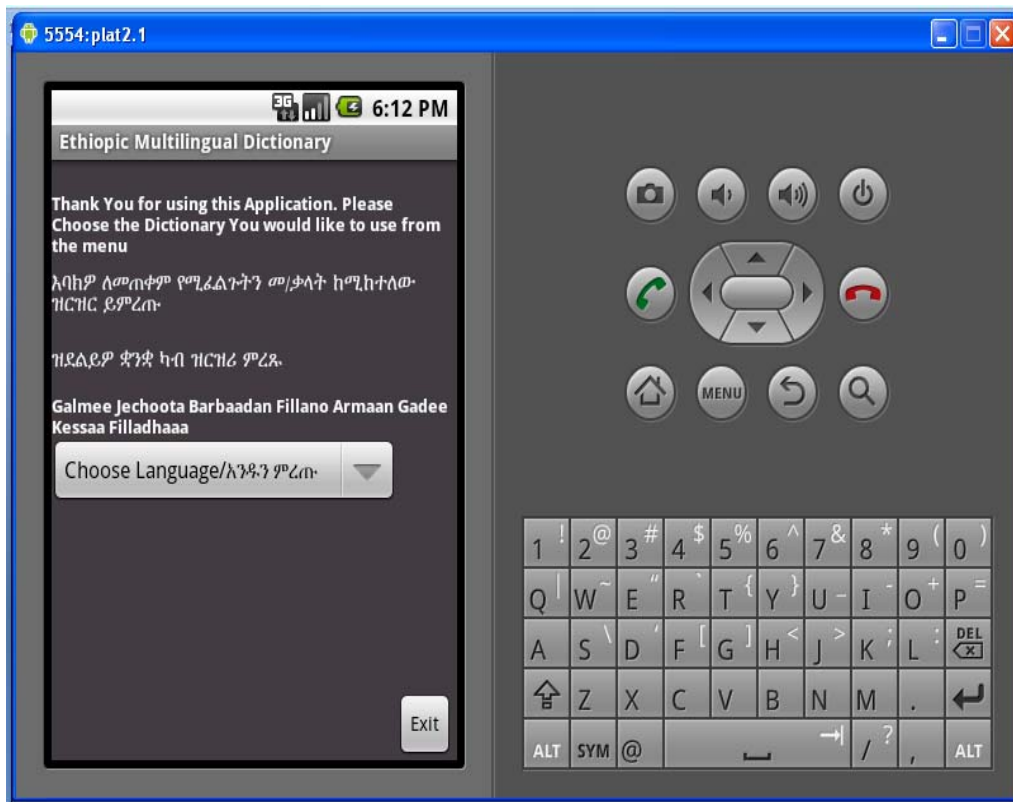


Figure 6-4 The startup screen for the Ethiopic Multilanguage dictionary application

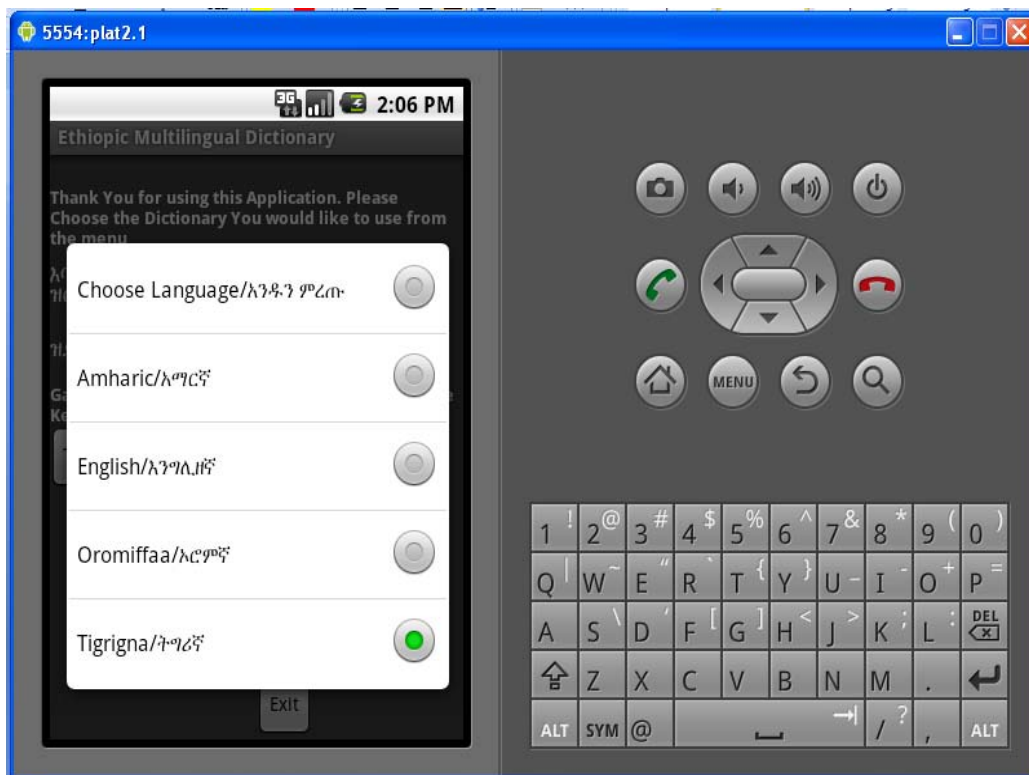


Figure 6-5 list of dictionaries included in the multilingual dictionary

The four activities are similar in structure except with the change in the interface display. Each activity has a textbox field to accept an input from the user. The user can use either the soft keypad or hard keypad to give an input. To use the soft keypad, the user needs to press down and hold onto the textbox for a while and the system displays the soft keypads available on the device. Then, the user selects the type of keypad to use from the Latin/Android keypad and Ethiopic keypad. On this project, the Ethiopic soft keypad designed and developed by [50] is used as an input method for the Ethiopic words and the default Android keypad is used for the Latin words. Depending on the chosen dictionary subtype, the user also needs to change the appropriate keypad type to use. At the moment the user starts typing a letter, the running activity fetches list of words that start with the given letter. The user can then pick a word from the list, if needed, or keep writing in which case the activity needs to filter out the word list displayed according to the input letters. The user can then choose the desired language into which the translation of the word is needed by clicking a button with the name of the translating language or can view the translation of the given word with all the translation languages at once. Finally the system displays the definition and pronunciation of the word in the chosen language. There is also an option for the user to jump from the current active dictionary subtype to another with a button click. Screenshots of some of the activities of the Ethiopic multilingual dictionary application are shown below.

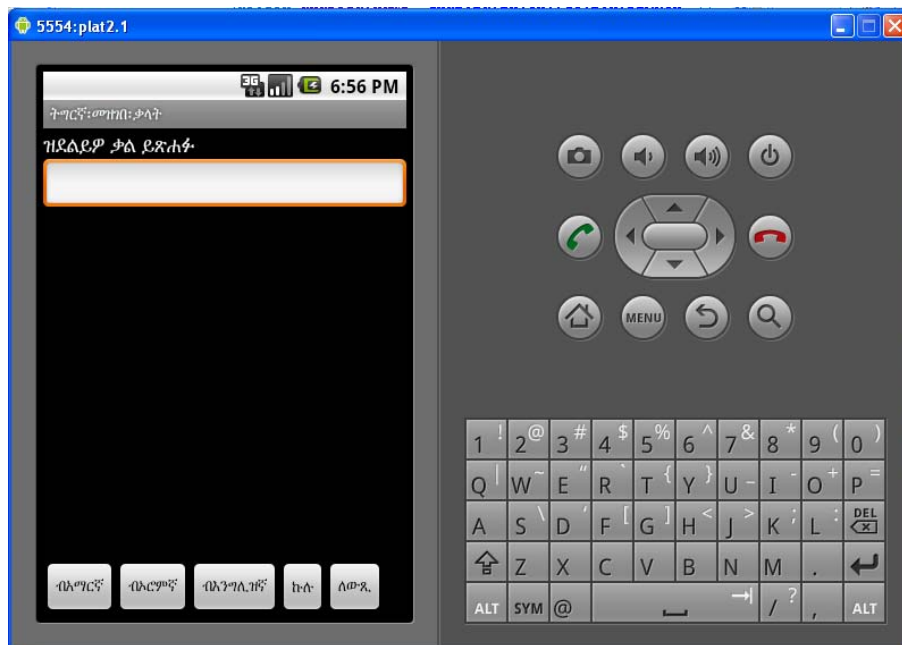


Figure 6-6 Screenshot for the Tigrigna Dictionaryentry subtype



Figure 6-7 List of soft keypads available on the device

As can be seen on Figure 6.7 above, there are list of soft keypads to choose from and use by the user. If the user wants to write English or Afaan Oromo words, he/she needs to make a choice on the Android keyboard keypad selection. For writing Amharic as well as Tigrigna words, the user is expected to choose the “የአማርኛ ኪቦርድ” keypad. After the user made the choice of keypads, the chosen one will be displayed on the currently active activity and is ready to be used as shown in Figure 6.8.

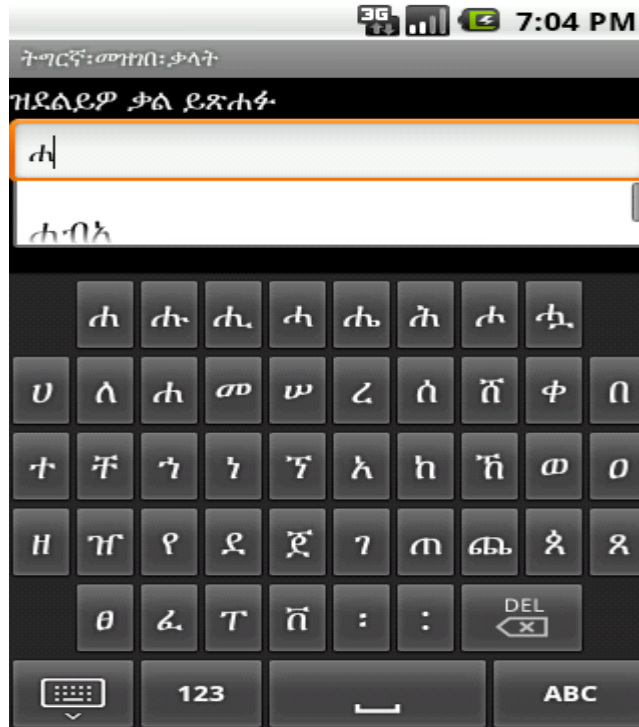


Figure 6-8 user selected keypad displayed on activity

The user then types the word to search for using the keypad displayed on the screen. The system waits for a single entry from the user and populates the screen with list of suggested words that might match the word needed to be translated to avoid the stress of the user for remembering the characters constituting a given word. Figure 6.9 is a screenshot taken from the system when displaying list of matching words as the user types the first letter on the Tigrigna dictionary entry.



Figure 6-9 suggested word list displayed

If the word the user is going to look for is included and displayed on the list, the user can choose that word and search for its translation and pronunciation in a required language. The user can also skip selecting the word from the list and type the whole word with the keypad.

To look for the translation and pronunciation of a given word, the user has to press one of the four labeled buttons on the activity. The first three buttons correspond to the language they have been labeled with. so for example, if a user typing on the Tigrigna dictionary subtype wants to see a translation and pronunciation of the typed word, say, in Amharic, then he/she is supposed to press the “ብኣማርኛ” button to display the corresponding result in Amharic language. The fourth button, if clicked displays all the translations and pronunciations of the given word in all the languages. A screenshot taken from the Tigrigna dictionary when the user selects one of the listed words and presses the “ብኣሮምኛ” button is shown below in Figure 6.10.

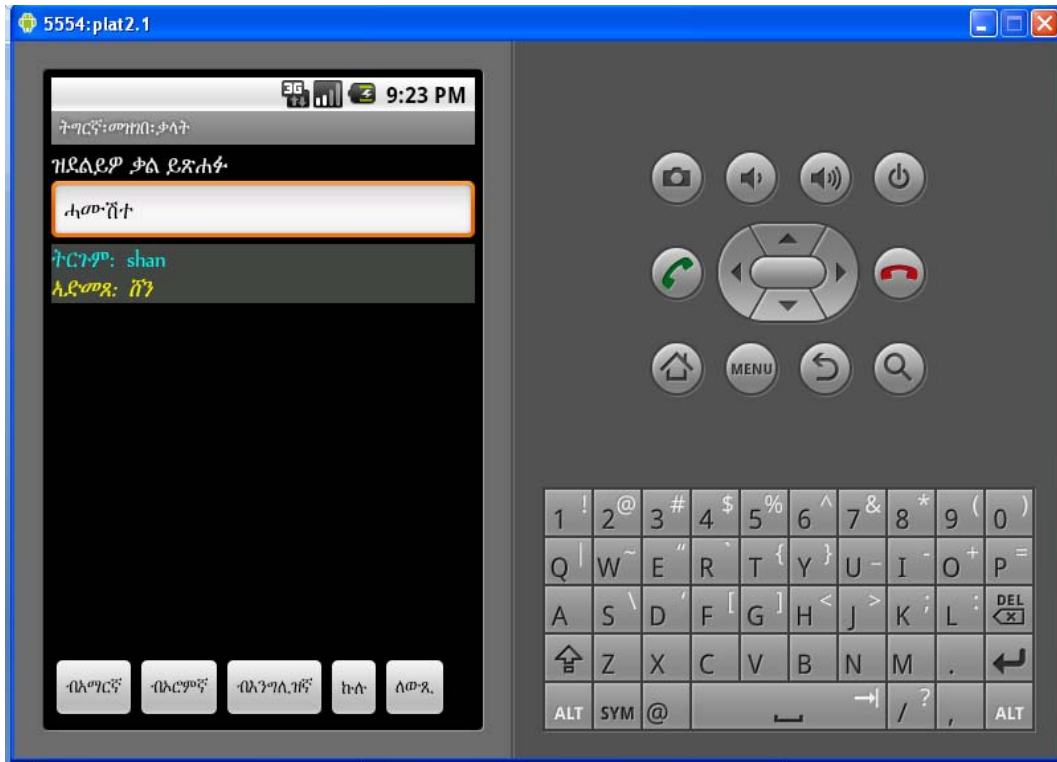


Figure 6-10 Translation and pronunciation of the Tigrigna word “ሓሙሽተ” in Afaan Oromo

Chapter Seven

Conclusion and Recommendation

7.1. Conclusion

The way people live, work, learn and communicate is shifting from paper to electronic form and nowadays even to mobile form. Recent mobile devices function both as a Personal computer and a phone at the same time. Smart phones are one of such devices which are capable of integrating and providing several user oriented functionalities in addition to making simple phone calls. As these devices are becoming more powerful and less costly, they are getting more user acceptance and are penetrating into every part of the world quickly. These devices use a proprietary Operating System since recently where the first open source mobile platform, Android, is released. This open source platform allows different developers to design and integrate several applications onto the devices. Moreover, the platform is designed to support different world languages in addition to English, which avoids the barriers encountered in designing and developing localized applications for any mobile device.

As a result, several language specific applications like EDs are being developed to be used on these devices that can be used as a language reference material. Designing and implementing an ED application that fits onto these devices is very crucial for any language learning as it can be consulted at the moment the user needs it with out the need of extra resources like computers, Internet connections and the like.

As one part of the world, in Ethiopia also the number of smart phone users is showing progress and increasing from day to day despite the fact that there are no applications embedded on the devices with local context. Providing some local applications on these devices would much more increase the number of users and also let the users interact with it using the languages they know.

Designing and implementing mobile multilingual ED application of the most broadly spoken and used languages in Ethiopia with some of the most widely used International languages like English, would be necessary as it provides an all-in-one reference with minimal cost. It can be used everywhere and anytime and also help users to learn those languages by themselves and communicate with each other easily.

In this work, a multilingual and multidirectional ED for the Amharic, Afaan Oromo, Tigrigna and English languages is designed and implemented on smart phones with Android platform. The work designed an appropriate UI that helps users to interact with it through the language they know and understand. A simple algorithm has been proposed that enables the system to fetch suggestion words from the persistent data store and that are expected to be what the user is looking for. Moreover, a general architecture for the ED had been designed and presented. The developed system is a multidirectional dictionary in that a words' meaning can be find in three languages and the three languages also have their translations in the other language.

This work has a great contribution towards facilitating smooth information exchange among people of the country, and also enhancing the bond between the local languages and the current communication technology. It can be of great help for language learners, international organization workers, tourists, investors, researchers, etc that must know the languages for the sake of understanding and communicating easily with the native speakers of the languages. Overall, as Ethiopia is a country with several ethnic groups and diverse language set, such a system plays a vital role towards aiding the development strategy the country is applying.

The system has been successfully implemented and tested on an emulator of Android based smart phones with a dictionary database of size 100KB. From the test, it has been observed that the system satisfies all the requirements expected from it by providing the desired output with negligible time delay. Moreover, users' opinion on the application has been also gathered to assure the importance of the system and its ease of use. Almost all respondents agreed on the simplicity of the interface designed and on the usefulness of the application especially being found on such mobile devices.

7.2. Recommendation

This is a first attempt to design and implement a Multilanguage ED of the Ethiopian languages in mobile devices which makes it need more refinement and improvement.

As future works:

- We propose extending this multilingual dictionary to talking dictionary by playing back the pronunciations for the words through sound rather than providing it as a simple text. Moreover it can be also made to accept inputs as sound rather than a text and display

back their definition and pronunciation through sound. This feature will make the application being used also by those who can not read and write texts like illiterates and the disabled.

- Extending this four-in-one multilingual dictionary to support more local as well as international languages without violating the resource constraints posed by the mobile devices is also another improvement that can be carried out on this project.
- Adding more functionalities like stemming and enabling the user to search for other derived forms of a given query would be an essential refinement to this work
- Letting the user to add more words on the dictionary database. If a word the user is looking for could not be found in the database, enabling the system to store those words as new records in the database would be another direction to enhance the developed system
- Integrating this dictionary application with other applications of the devices and making it to be used as a reference material with in these applications can also be one direction to increase the usage of the application.

References

- [1]: Xiaohua Du and Xiaoyi Sheng, “Smart phone”, <http://www.cis.upenn.edu/~palsetia/cit595s08/Projects08/Smart-Phone.ppt>, last accessed on February 03, 2010
- [2]: electronics information online, “Cellular Phone”, <http://www.electronics-manufacturers.com/info/communication-equipment/cellular-phone.htm/> , last accessed on February 03, 2010
- [3]: Malcolm Tatum , “What is a Smart phone?”, <http://www.wisegeek.com/what-is-a-smart-phone.htm/> , last accessed on February 03, 2010
- [4]: Gartner 2009 Press Release: <http://www.gartner.com/it/page.jsp?id=1256113>, last accessed on February 05, 2010.
- [5]: <http://code.google.com/android/>: Last accessed on January 29, 2010
- [6]: “Android (Smart phone)”, [http://en.citizendium.org/wiki/Android_\(smart_phone\)](http://en.citizendium.org/wiki/Android_(smart_phone)) last accessed on February 10, 2010
- [7]: “Electronic dictionary”, http://en.wikipedia.org/wiki/Electronic_dictionary/, last accessed on March 10, 2010
- [8]: Robert Weschler and Chris Pitts, “ An Experiment Using Electronic Dictionaries with EFL Students”, <http://iteslj.org/Articles/Weschler-ElectroDict.html/>, last accessed on March 10, 2010
- [9]: Huckin, T., Haynes, M., & Coady, J., 1993, “Second Language Reading and Vocabulary Learning”, Norwood, NJ: Ablex.
- [10]: Sobah Abbas Petersen, Jan-Kristian Markiewicz, 2008, “PALLAS: Personalized Language Learning on Mobile Devices, Fifth IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education
- [11]: <http://www.aau.edu.et/index.php/academics/ethiopian-languages-research-center>, last accessed on March 28, 2010
- [12]: Michael Zock, Slaven Bilac, 2004, “Word lookup on the basis of associations: from an idea to a roadmap”, ACL Workshops Proceedings of the Workshop on “Enhancing and Using Electronic Dictionaries”, Geneva, Switzerland
- [13]: Mary Elizabeth, “What is a Dictionary?”, <http://www.wisegeek.com/what-is-a-dictionary.htm/> last accessed on May 5, 2010

- [14]: Concise Amharic Dictionary [Paperback], <http://www.amazon.com/Concise-Amharic-Dictionary-Wolf-eslau/dp/0520205014/>, last accessed on March 5, 2010
- [15]: Amsalu Aklilu ,1979 EC, “*አማርኛ-ፎንዳኒኛ መዝገበ ቃላት* Amharic-English Dictionary”, Kuraz Publishing Agency.
- [16]: Mahdi, Hamid Muudee. 1995. “*Oromo Dictionary, English-Oromo*”, Vol.I, Atlanta, Georgia: Sagalee Oromo Publishing co.
- [17]: Tamene Bitima. 2000. “*A Dictionary of Oromo Technical Terms: Oromo-English.*” GTZ Primary Education Assistance Project (PEAP) with the Oromia Regional Education Bureau (OREB), Finfinnee.
- [18]: Daniel Liden, “What Is an Electronic Dictionary?”, <http://www.wisegeek.com/what-is-an-electronic-dictionary.htm>, last accessed on May 5, 2010
- [19]: Lehr, A. 1996. ‘Zur neuen Lexicographica-Rubrik, ‘Electronic dictionaries’’. *Lexicographica* 12: 310–7.
- [20]: Hayward, Katrina and Richard J. Hayward. 1999. “Amharic. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*”, Cambridge: the University Press.
- [21]: “Amharic Language”, http://en.wikipedia.org/wiki/Amharic_language, last accessed on May 5, 2010
- [22]: Daniel Yacob. 2005, “Developments Towards an Electronic Amharic Corpus”, White paper
- [23]: Yibeltal Tafere, 2008, “FORMANT-BASED SPEECH SYNTHESIS: A CASE OF AMHARIC WORDS”, AAU, School of Graduate Studies
- [24]: Wolf Leslau. 1969, ”An Amharic Reference Grammar”, California University, Dep’t of Near Eastern and African Languages, USA
- [25]: Baye Yimam. The Phrase Structure of Ethiopian Oromo. In *Ph.D. Thesis*. School of Oriental and African Studies, University of London, 1986.
- [26]: K. Kekeba Tune, V. Varma and P. Pingali, “Evaluation of Oromo-English Cross-Language Information Retrieval
- [27]: “Oromo Language”, http://en.wikipedia.org/wiki/Oromo_language, last accessed on March 12, 2010
- [28]: Kebede Bekere, “500 Word Afaan Oromifa Dictionary”, <http://www.shenpres.org/FellowshipICare/Oromifa.htm#background>, last accessed on March 12, 2010

- [29]: “Tigrigna Language”, http://en.wikipedia.org/wiki/Tigrigna_language, last accessed on March 12, 2010
- [30]: P. Bernard, J. Dendien and J. Pierrel, “A Computerized Dictionary: Le trésor de la langue française informatisé (TLFi)”, white paper accessed from <http://acl.ldc.upenn.edu/W/W04/W04-2107.pdf>
- [31]: project note: G. Neumann, C. Fellbaum, A. Geyken, A. Herold, C. Hümmer, F. Körner, U. Kramer, K. Krell, A., Sokirko, D. Stantcheva and E. Stathi, “A Corpus-based Lexical Resource of German Idioms”, 2004, Proceedings of the ACL Workshop on Enhancing and Using Electronic Dictionaries, Geneva, Switzerland
- [32]: Igor A. Bolshakov and Alexander Gelbukh; “A Very Large Dictionary with Paradigmatic, Syntagmatic, and Paronymic Links between Entries”, COLING 2004, The 20th International Conference on Computational Linguistics
- [33]: James McCracken. “Rebuilding the *Oxford Dictionary of English* as a semantic network”, white paper accessed from <http://acl.ldc.upenn.edu/coling2004/W10/ps/14.ps>
- [34]: M. Schmidt, S.Fitt, C. Scott and M. Jack. “Phonetic Transcription Standards for European Names (ONOMASTICA)”, 1993, http://www.cstr.ed.ac.uk/publications/new/1993/Schmidt_1993_a.ps
- [35]: Vlado Kešelj, Tanja Kešelj, Larisa Zlatić. “English—Serbo-Croatian Electronic Dictionary”: white paper accessed from <http://aclweb.org/anthology/w/w04/w04-2112.pdf>.
- [36]: Svetlana Sheremetyeva, 2004, “Application Adaptive Electronic Dictionary with Intelligent Interface”: in proceeding of the workshop on Enhancing and Using Electronic Dictionaries, Geneva Switzerland.
- [37]: Gdbór Prószóky. 1998. “An Intelligent Multi-Dictionary Environment”: in proceeding of the 17th International Conference on Computational Linguistics-Volume2 Montreal, Quebec, Canada.
- [38]: Gábor PRÓSZÉKY, Balázs KIS, 2002, “Context-Sensitive Electronic Dictionaries”: in Proceeding of the 19th international conference on computational linguistics-volume2 Taipei, Taiwan.
- [39]: M. Tomokiyo, M. Mangeot, E. Planas, N. Dessaigne, Geta-clips-imag. “Papillon: a Project of Lexical Database for English, French and Japanese, using Interlingua Links” White paper
- [40]: Paragon Software. Multi-language dictionary “SlovoEd” for NOKIA Communicator
- [41]: “ABBY Launches Comprehensive Multilingual Dictionary for Mobile Devices”, <http://www.abby.com/Default.aspx?DN=5d823ebe-44fd-47e3-a143-aa4a9cdef31>, last accessed on March 24, 2010

- [42]: <http://www.amharicdictionary.com/2.0/Default.aspx>, last accessed on May 26, 2010.
- [43]: <http://www.memhr.org/dic/>, last accessed on May 20, 2010
- [44]: M. Fowler, K. Scott and G. Booch, 1999, “UML Distilled: A Brief Guide to the Standard Object Modeling Language”
- [45]: Mark L. Murphy. 2009. “Beginning Android” Apress and friends ED books, Berkley, CA
- [46]: Hinsene Mekuria (M.A). 2010 “The Genius English-Amharic Dictionary” Commercial Printing Enterprise
- [47]: Hinsene Mekuria. 2006 “DUNGOO; Oromo-Amharic-English Dictionary” Commercial Printing Enterprise
- [48]: Efreem Zekarias “Tigrigna-English, English-Tigrigna Dictionary glossary” accessed from www.memhr.org May 15, 2010
- [49]: Reto Meier. 2009. “Professional Android™ Application Development”, Wiley Publishing, Inc., Indianapolis, Indiana
- [50]: Andualem W., Yonnan G., Gedion T., Solomon A., “The “ሀለክ” Virtual Ethiopic Keyboard for Smart Phones”, Addis Ababa University, Department of Computer Science, unpublished article

Declaration

I, the undersigned, declare that this project is my original work and has not been presented for degree in any other university, and that all sources of materials used for the project have been duly acknowledged.

Declared by:

Name: Amsale Zelalem Bayih

Signature: _____

Date: June 14, 2010

Confirmed by advisor:

Name: _____

Signature: _____

Date: _____

Place and date of submission: Addis Ababa, June 2010.