ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

COREFERENCE RESOLUTION FOR AMHARIC TEXT USING BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMER

Lingerew Bantie Asmare

A thesis submitted to the department of Computer Science in partial fulfilment for the degree of Masters of Science in Computer Science

Addis Ababa, Ethiopia

March 4, 2022

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

Lingerew Bantie Asmare

Advisor: *Yaregal Assabie (PhD)*

This is to certify that the thesis prepared by *Lingerew Bantie*, titled: c*oreference resolution for Amharic text using bidirectional encoder representation from transformer* and submitted in partial fulfilment of the requirements for the Degree of Master of Science in Computer Science complies with the regulations of the university and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name	Signature	Date
Advisor: Yaregal Assabie (PhD)	__________________	_______________
Examiner: Fekade Getahun (PhD)	__________________	_______________
Examiner: Solomon Gizaw (PhD)	__________________	_______________

Abstract

Coreference resolution is the process of finding an entity which is refers to the same entity in a text. In coreference resolution similar entities are mention. The task of coreference resolution is clustering all similar mentions in a text based on the index of a word. Coreference resolution is used for several Natural Language Processing (NLP) applications like machine translation, information extraction, name entity recognition, question answering and others to increase their effectiveness. In this work, we have proposed coreference resolution for Amharic text using bidirectional encoder representation from transformer (BERT). This method is a contextual language model that generates the semantic vectors dynamically according to the context of the words.

The proposed system model has training and testing phase. The training phase includes preprocessing (cleaning, tokenization and sentence segmentation), word embedding, feature extraction Amharic vocabulary, entity and mention-pair and coref model. Like training phase, testing phase has its own step such as preprocessing (cleaning, tokenization and sentence segmentation) and coreference resolution as well as Amharic predicted mention. The use of word embedding in the proposed model is that it represent each word into a low dimension vector. It is a feature learning technique to obtain new features across domains for coreference resolution in Amharic text. Necessary informations are extracted from word embedding and processed data as well as Amharic characters. After we extract important features from training data we build a coreference model. Moreover, in the model bidirectional encoder representation from transformer is used to obtain basic features from embedding layer by extracting various information from both the left and right direction of the given word.

To evaluate the proposed model, we conduct the experiment using Amharic dataset, which is prepared from various reliable sources for this study. The commonly used evaluation metrics for coreference resolution task are MUC, B^3, CEAF-m, CEAF-e and BLANC. Experimental result demonstrate that the proposed model outperformed state-of-the-art Amharic model achieving 80%, 85.71%, 90.9%, 88.86% and 81.7% F-measure values respectively on the Amharic dataset.

Keywords: Amharic coreference resolution, mention, Bidirectional encoder representation from transformer, Transformer, NLP, Coreference, word embedding

Dedication

This work is dedicated to my father Bantie Asmare and my mother Emebet Wasie as well as my sister Ayal Akinaw.

Dad: You are the hero that you did to raise me in my life. I am so grateful for being your son.

MOM: You are the second hero for your long lasting love and care for me in my life history. You are the icon of all things.

Acknowledgement

First and foremost, I would like to thank the almighty God for his support in my life. Next to this, I would like to thank my precious advisor Dr. Yaregal Assabie for his energetic support, fantastic advice and assistance without feeling tired up to the end of this work. You commented me starting from title selection to the end of the thesis work. You are so special and most precious one.

I want to thank my family for their support in every situation in my life. I have never seen any person like you in my life. You are so hard, consistent and talent for everything. I also want to thank my friends Mr. Abreham Bekele and Mr. Mekonin Kasu for their help, appreciation and support from the beginning to the end of this thesis. I also would like to thank all my classmates for their encouragements and supports. Next to this I would like to thank Mr. Grima Kasahun from linguistic department for his willingness to give comments on linguistic aspect and dataset preparation.

Last but not least, I would like to thank Mr. Ysacor Mengistu for encouraging me to complete the thesis successfully.

Thank you so much!!!

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

AAiT -Addis Ababa institute of Technology

BERT- BERT Bidirectional Encoder Representations from Transformer

BiLST- Bidirectional Long Short Term Memory

BLANC- BiLateral Assessment of Noun-phrase Coreference

CEAF-e - Constrained Entity Aligned F-Measure entity

CEAF-m - Constrained Entity Aligned F-Measure mention

CONLL- Conference on Computational Natural Language Learning

CoNLL- Computational Natural Language Learning

CNN - Convolutional Neural Networks

FFNN - Feed Forward Neural Network

LSTM - Long Short Term Memory

LEA- Link-Based Entity-Aware

MELA- Mention, Entity, and Link Average score

MUC- Message Understanding Conference

NER- Name Entity Recognition

NLTK- Natural Language Toolkit

NLP- Natural Language Processing

NP- Noun Phrase

NPs- Noun Phrases

PoS- Parts of Speech

PP- Personal Pronoun

RAM- Random Access Memory

RNN - Recurrent Neural Networks

SVM- Support Vector Machine

CHAPTER ONE: INTRODUCTION

1.1 Background

Nowadays researchers are increase due to the increasing of applications in various technologies. To do those applications researchers uses Natural Language Processing application. Natural Language Processing (NLP) is the process of analysing human language text by using computerized approach that is based on a set of technologies. It is used for studying and representing naturally existing text [12]. There are a number of unstructured data available on the Internet. This abundant data is useful only if suitable techniques are applied to process the data and identify key terms from it. This process needs natural language processing application; from that coreference resolution is one of such significant applications [12].

Coreference resolution is the process of finding all key terms that are referring to the same entity in a sentence [3, 4, 5, 6, 14, 23]. Apart from that, coreference resolution is the duty of collecting all mentions of entities in a sentence into same classes so that all the mentions in a given class refer to the same entity [3, 6]. Assuming that the referring expressions α_1 and α_2 are occurrences of noun phrases, and that both have a unique reference in the context in which they occur, Van Deemter and Kibble [47] state that α_1 and α_2 co-refer if and only if reference (a_1) exactly equal to reference (a_2). The main task of coreference resolution is to band the mentions and place them together into identical classes that are referring to the same entity in a sentence. Investigation of the relation that existed between two entities is a complex task in a sentence; how they are related and which entity is referring to which entity in sentences is not easy to determine[22, 24, 28, 33, 40].

Resolution of coreference is one of the most challenging and big tasks in the field of Natural Language Processing [5]. In order to understand the coreference resolution task in detail, first we have to know two basic words: *Entity* and *Mention*. An entity is a set of objects that occur in the real world to represent something [6]. Moreover, it is a blend of textual references to the similar object in the sentence. A mention is an individual textual reference of an entity that referring to similar expression in a document [5].

According to Mitkov *et al.* [3], the task of coreference is usually complicated with the task of anaphora resolution. The word coreference and anaphora are occasionally used interchangeably, but they are not always the same. In anaphoric resolution the reader goes back in the text to find their antecedent [9]. But a coreference resolution could be stand on its own which means that they have an equivalence relation [7].

Sometimes coreferent expression will be anaphoric. When we say a co-referent expression is anaphoric, if its interpretation is depending on a previous expression in the sentence [7]. Generally, the task of coreference resolution is different from anaphoric expression because some coreferent terms are not anaphoric terms. Let us take an example to show their difference.

Example 1: *Captain Abebe is a good seaman, worthy of the ship he commanded. His vessel and he were one. He was the soul of it.*

From this example *{the ship, his vessel, and it}* are coreferents but *{his vessel, it, he}* are anaphoric terms. Therefore, some expressions are coreferential but not anaphoric.

Example 2: *A bus had to divert to the local hospital when one of the passengers had a heart attack. It go to the hospital in time and the man's life was saved.*

From Example 2 *{the local hospital, the hospital}, {bus, it}* and *{one of the passengers, the man}* are coreferential expressions [40]. Whereas *{it}* is anaphoric term. Accordingly, some expressions are both coreferential and anaphoric.

To get clear and more understanding about coreference resolution look at the following examples.
Example 1: *John lives in Addis Ababa. He is quite happy in that city.*

John: *{John, He}* and Addis Ababa: *{Addis Ababa, that city}*

All the entities are "*John*" and "*Addis Ababa*" with their referents "*He*" and "*that city*" respectively. Therefore, "*He*" and "*John*" are coreferent terms and also "*Addis Ababa*" and "*that city*" are coreferents.

Example 2: *Mary has a pen. John loves her. She was born in 1993. He lives in Ethiopia.*

Coreferential: - *{Mary, her, She}* and *{John, He}*

Example 3: *Barack Hussen Obama, the former American president, has told the country he is ready for a long vacation*.

Coreferential: - {*Barack Hussen Obama, the former American president, he}*

Therefore, a referring expression (mention) could be noun phrases (*John*), named entity (*Addis Ababa*), or pronoun (*He*), which is refers to an entity in the real world. A collection of referring expressions with the same referent is said to be a coreference chain or cluster. The main objective of a coreference resolution system is to produce all the coreference chain of a given sentence. See the above three separate sentences with their corresponding coreference chain [16].

Most of the researches carried out in the area of coreference resolution were intended to the resolution of proper noun and pronominal using rule-based approach because it is very complex task to deal with the complete coreference resolution. There are various approaches that are applied to develop coreference resolution effectively. From that rule based, machine learning [41], transformer based [16] and hybrid [39] as well as deep learning approaches are used [32]. Many researches use rule-based and machine learning approach for coreference resolution [45]. Some method (rule-based) is lack of portability and robustness. To address these problems, transformer based approach (Bidirectional Encoder Representation from Transformer (BERT)) is proposed. BERT language model requires language-specific manually annotated data for training and testing purpose [23]. This annotation process is fastidious and needs human labour and time. Supervised machine learning technique in [54] developed for this task based on a sequence labelling such as Hidden Markov Models [9], conditional random fields and machine learning classifiers based such as support vector machines (SVM) [4], decision tree and so on. In contrast, semi-supervised coreference resolution is used both label and unlabelled text which expected to decrease the labour intensive and the cost of developing such system [11]. By considering the massive unlabelled data freely available to learn word representation and use it as features to boost supervised coreference resolution system. Hybrid approach [39] use the synergy of rule based and machine learning approaches [52]. Transformer based model is a deeply learning methods, that encode each words accurately to find the meaning of each word with its context.

A number of works have been done to coreference resolution for English and other languages. In Ethiopian language, limited work is found only for Amharic language and other Ethiopian languages [11]. Thus, the main objective of this research is to develop coreference resolution for Amharic text using BERT that increases system efficiency and effectiveness.

1.2 Motivation

There are over 80 languages spoken in Ethiopia. However, Amharic language is the working language of the federal government of Ethiopia with the population of over 100 million. Amharic language has its own script and alphabet. This language has many usages in different sector of Ethiopia. Because of its importance, there has been a steady growth of research and development on Amharic natural language processing [11]. Due to the ongoing expansion of the research, availability of Amharic language research is dramatically increased in day-to-day life. From that research in Amharic coreference resolution is one of the key important components in natural language processing application. Amharic noun phrase coreference resolver is used to determine whether two noun phrases are referring to the same real-world entity or not. The resolution of Amharic coreference is complex it needs to be addressed strongly for the effectiveness of NLP applications like information retrieval, information extraction, question answering, machine translations and many others [16]. Amharic coreference resolution is an interesting thematic area to find coreference expression in the text. This Amharic coreference resolution is used for various NLP applications such as information extraction, information retrieval, question answering, automatic abstracting, and machine translation and so on.

In information extraction, coreference resolution is required in order to group information about the same entity or referent that is located in different parts of the discourse. For information retrieval if users want to retrieve documents from different corpus, it will generate the intended document based on his/her query. So, the query is answered by an entity from the document [16]. In question answering system coreference resolution is used to answer the question. For automatic summarization and automatic abstracting, it is used to improve techniques for extracting most relevant and related sentences. In machine translation system coreference resolution is important to resolve pronouns to their antecedents when translating from one language in to others [2]. The use of coreference resolution in various applications improves their performance effectively [2]. This has motivated us to do research on Amharic coreference resolution using transformer based model.

1.3 Statement of the Problem

Coreference resolution has been studied for different languages (English, Swedish, German, Arabic, polish, French, Russia etc.) using various techniques[18]. Wallin and Nugues [21] proposed coreference resolution for Swedish and German using distant supervision. Beseiso and Al- Alwani [13] also studied coreference resolution for Arabic text using morphological features. In addition to this, Niton *et al.* [19] developed coreference resolution for Polish using deep neural network. Moreover, Toldova1 and Ionov [20] proposed mention detection for improving coreference resolution in Russian text using machine learning approach. Zhekova [55, 20] also proposed multilingual coreference resolution for English text using machine learning approach. Hai-Long *et al.* [23] proposed coreference resolution for English full text articles using BERT. The system achieved high performance using BERT over the baseline model of LSTM. Significantly increase the performance of the system in both mention detection and coreference resolution. BERT allows to learn the context of a word-based on all of its left and right of the masked word. This language model is most important to generate the context of the masked word.

Review of related works reveals that coreference depends on language specific features. Thus, researches proposed various techniques and algorithms based on the unique characteristics of the specific languages. Accordingly, Temesgen Dawit and Yaregal Assabie [11] proposed Amharic anaphora resolution system using knowledge-poor approach by considering the unique morphological features of the language. They identify Amharic dependant pronoun on verbs and independent pronoun for anaphora resolution. They tested their implementation with the critical success rate of value 81.79% for hidden anaphora and 70.91% for independent anaphora. However, coreference resolution is a super set of anaphoric resolution and thus not all coreferences are considered in the study. Thus, in this work we use BERT to develop Amharic coreference resolution.

1.4 Objectives

General Objective

The general objective of this study is to design and develop coreference resolution for Amharic text using bidirectional encoder representation from transformer.

Specific Objective

To conduct different literature reviews and related work in the area of coreference resolution the following specific objectives are considered.

- To conduct different literature reviews and related work in the area of coreference resolution as well as different approaches
- To collect Amharic documents to be used as dataset
- To study the nature and characteristics of Amharic grammar and coreferent terms
- To develop a model for Amharic coreference resolution
- To evaluate the performance of the system.

1.5 Methodology

The methods followed to achieve the objectives of this work are described in the following sections.

Literature Review

Related literatures are reviewed from different sources (books, journals) to understand the nature of coreference resolution in particularly most important to Amharic coreference resolution. We have reviewed various kinds of existing approaches to solve coreference resolution problem. In addition to this, the existing designed coreference resolution system for other languages is reviewed. Finally, to understand the nature and structure of Amharic grammar, various books, articles, newspapers, holy bibles and others, which are useful for this is reviewed.

Data Collection

We grasp Amharic text from various reliable sources for this study. These documents are collected and organized from multiple sources which have enough coreferent terms. Therefore, the reliable sources are from books, encyclopaedias, fictions, bibles, Wikipedia, and so on. The collected documents are used for training the model and testing the system after we made pre-processing step.

Tool

We used python programming language to develop a prototype for Amharic coreference resolution with the help of various libraries.

Prototype Development

We developed prototype of the system to test and train the system. The prototype is used to test the performance of the system. We used Python programming language to build a coref model of the system.

Evaluation

The performance of the proposed system is evaluated based on the standard metrics of coreference resolution score, which are expressed in terms of Precision, Recall and F-measure. For evaluating the performance, it should have different documents.

Testing

After evaluating the performance of the system by various metrics, the system will be tested to check the accuracy measure of the system with Amharic text.

1.6 Scope and Limitations

The scope of this thesis is to develop coreference resolution for Amharic text. Its scope is limited to process only machine editable text. Scope of the research work is also limited in identifying coreferent terms from Amharic sentence.

The Amharic coreference resolution model is applied only for machine editable text. In this work we have extracted mentions, but not include fully identification of singleton.

1.7 Application of Results

Coreference resolution is one of a major component of natural language processing. It has many potential downstream applications in natural language processing like machine translation, paraphrase detection, summarization, question answering, full text understanding etc. Therefore, Amharic coreference resolution will be used for Amharic machine translation, Amharic text summarization, and named entity recognition in Amharic documents and for Amharic question answering system.

In question answering system, coreference resolution is used to understand the exact interpretation of a text, or to identify the relative significance of peculiar subject, pronouns and other referring expressions must be linked to the right person or entity. For full text understanding, coreference resolution is used to comprehend the full message of the given sentence accurately with in a limited time. Therefore, at the end of this research, students (other researchers) will use it, to do a full-fledged application from the above stated applications.

1.8 Organization of the Rest of the Thesis

This paper talks about Amharic coreference resolution that contains different chapters. These chapters of the thesis are organized as follow. Chapter two discusses briefly the process of coreference resolution, the models of coreference resolution and its approaches, concepts about Amharic language and other related topics which are necessary to it. The third chapter contains various works which are related to this topic but there is no well-developed work in the area of Amharic coreference resolution. Next to that, chapter four presents design and implementation of new system as well as the architecture of the proposed system (architecture of Amharic coreference resolution). The fifth chapter is about implementation and evaluating the performance of the system. The last chapter is conclusion and recommendation.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

In this chapter some of the overview of the most significant theoretical knowledge about the Amharic language are discussed. The sections of this chapter deals about Amharic writing system, Amharic morphology, coreference in Amharic and coreference resolution with various techniques as well as standard evaluation metrics to coreference research work.

2.2 Amharic Language

Amharic language is the working language of the federal government of Ethiopia with a population of over 110 million. In this research Amharic text is considered. It is important to investigate the characteristics/features of the language. Amharic is one of the Ethiopian semitic languages, which are sub-grouping under the Afro-Asiatic languages. Amharic language is spoken as a first language by the people of Amhara and others that serves as the working/official language for business and administrative sectors in all cities and towns. Languages are not a simple tool for communication but it offers an insight into a different culture and way of living. Amharic language is classified under semitic language family that could be a national language of African country. It is also an official language of the federal democratic republic of Ethiopia. It is the second most voice communication among semitic language families next to Arabic. Amharic language has its own writing script that originated from the Ge'ez alphabet. The whole characters of Amharic language are *fidels*. Amharic *fidel* has at least forty one consonant and seven vowels. It is written from left to right style. Amharic writing style and Amharic morphology are discussed in the following sub-sections [1, 11].

2.2.1 Amharic Writing System

An Amharic language has its own writing system that is Fidel, which is adapted from Ge’ez language. It has been used characters from Ge'ez, which has redundant and non- redundant characters in the writing system. For instance, the redundant characters ሀ, ሐ and ኀ are pronounced as (hä), ሰ and ሠ are pronounced as (sä), and አ and ዐ are pronounced as (a), ጸ and ፀ are pronounced as (tsä). Amharic non-redundant characters are ለ, መ, ረ, ጠ, ነ, ፈ, ወ, ደ, የ, ከ, ገ, ተ, በ and so on. Amharic alphabet does not have upper-case and lower-case in nature. It has a total of 435 characters [59].

Amharic language has its own punctuation marks. There are varieties of punctuation marks used in Amharic writing system. Among them some of the most common punctuation marks are four dot (::), comma (፣), colon (፥), question mark (?), exclamation mark (!) and so on. Most of the time, four dot (::), question mark (?) and exclamation mark (!) are used to sentence demarcation [60].

2.2.2 Amharic Morphology

Amharic language is one of the most challenging, morphologically rich and complex Ethiopian language. In this language, different affixes are used to make inflectional and derivational morpheme. The derivation is achieved by affixation (prefix, infix, and suffix) or compounding. Unlike English language gender, number, prepositions and others information unit of measurement attached to Amharic nouns and adjectives that resulted in advanced morphology of the language. For instance, from the noun በግ/sheep the following word unit of measurement generated through inflection and affixation. These are በጎች/sheeps, በጉ/ the sheep, በጌ/my sheep, በጌን/my sheep, በግ/your sheep, በግህ/ your sheep and so on. It is to boot potential to come back up with the following words from the adjective ትልቅ/big. These are ትልቁ/ very big, ትልቆች/very big, ትልቆቹ/very big, and so on. Inflections and derivations of Amharic verb unit of measurement even a great deal of advanced than that of Amharic nouns and adjectives. It is as a result of many verbs in surface form unit of measurement that is generated from one verbal stem and much of stems in turn unit of measurement generated from one verbal root. Combination of person, gender, number, case, tense/aspect, and different information unit of measurement extracted from Amharic verbs resulting in thousands of words from one verbal root. As a result, one word might represent a complete sentence constructed with subject, verb and object [11, 61].

2.2.3 Coreference in Amharic

Amharic is the working language of the the federal government of Ethiopia. This language needs improvement to process various document. To the best of our knowledge, coreference resolution for the Amharic language using bidirectional encoder representation from transformer is not done so far. In this thesis, the coreference resolution model works based on Amharic sentence structure. Amharic language has different structure and syntax [1]. The given example below shows that the labeled sentence for Amharic language. The types of coreference link are discussed using Amharic wikipedia text as followed.

<COREF ID="0" TYPE="IDENT">ኃይሌ ገብረ ሥላሴ </COREF>ሚያዝያ 10 ቀን 1965 ዓ.ም <COREF ID="1" TYPE="IDENT">በአሰላ ከተማ</COREF> አርሲ ክፍለ ሀገር ተወለደ። <COREF ID="0" TYPE="IDENT">ሯጭ ኃይሌ</COREF> የሚኖረው <COREF ID="2" TYPE="IDENT"> ኢትዮጵያ </COREF> ውስጥ ነው። ኃይሌ <COREF ID="2" TYPE="IDENT">አገሩን</COREF > በጣም ይወዳል። <COREF ID="0" TYPE="IDENT">ኃይሌ ገብረ ሥላሴ</COREF> እቻ ያልተገኘለት ድንቅ የረጅም ርቀት <COREF ID="0" TYPE="IDENT">ኢትዮጵያዊ ሯጭ</COREF> ነው። <COREF ID="0" TYPE="IDENT"> ኃይሌ </COREF> በሩጫ ዘመኑ በ 10 ሺህ, በ 5 ሺህ በግማሽ ማራቶን, በማራቶንና እንዲሁም በሌሎቹ የሩጫ ዓይነቶች ከ 11 በላይ የዓለም ክብረ-ወሰን ሰብሯል። <COREF ID="0" TYPE="IDENT"> እሱም </COREF>በሩጫ ችሎታው ጠንካራ የሆነበት ምናልባትም <COREF ID="1" TYPE="IDENT">

ከመንደሩ </COREF> ከባሕር ጠለል በላይ ከፍታ ነው ተብሎ ቢታመንም የተፈጥሮ ጥንካሬው ዓለምን ያስደነቀ ብርቅዬ <COREF ID="0" TYPE="IDENT">ኢትዮጵያዊ</COREF> ነው።

From this we conclude that {ኃይሌ ገብረ ሥላሴ, ሯጭ ኃይሌ, ኃይሌ, እሱም} are coreferents. Also the second coreferent terms that is extracted from the given sentence are {ኢትዮጵያ and አገሩን} and {አሰላ ከተማ and መንደሩ}. As we can see from the above example, coreference resolution is a harder task to identify all mention in the sentence. Therefore, the task needs to know the characteristics and properties of a language. For instance let us take number agreement in Amharic language.

Example 1: *አበበ አዲስ መኪና አለው። እሱም ቀይ ነው።*

Example 2: *አበበ ሦስት አዳዲስ መኪናዎች አሉት። እነሱም ቀይ ናቸው።*

From the first sentence "አዲስ መኪና" and "እሱም" agree in number and from the second sentence *"ሦስት አዳዲስ መኪናዎች"* and *"እነሱም"* are also agree in number.

2.3 Coreference Resolution

Coreference resolution is the process of identifying all noun phrases that refer to the same real world entity in text. The task of reference resolution is to determine which noun phrases refer to each real world entity mentioned in the text. Users need coreference resolver to find out all noun phrases (NPs) easily that are referring to the same real world entity in a sentence [31, 32, 33]. In coreference resolution noun phrases can be a pronominal phrase (he), a nominal phrase (the queen)...etc. There are several data available online to be processed using NLP applications. These data are present without any sequence of text and it needs to apply some application on it in the context of NLP. During this time discourse plays an important role to identify the order of sentence that appear one after the other. In the discourse there will be an entity and possible references to those entity, obviously it is called "mention" [31].

Example: *Genet is a clever student in AAiT. She likes working research in natural language processing at the institute of Engineering. Her hobbies include reading, watching films and helping others.*

"*Genet*", "*AAiT*", and "*natural language processing*" are possible entities. "*Genet*" is the possible antecedent of pronoun "*Her*" and "*She*". So, "*She*" and "*Her*" are references to the entity "*Genet*" and "*the institute of engineering*" is a reference to the entity "*AAiT*". In the context of natural language processing, reference is a linguistic process where one word in a discourse may refer to another word or entity. The task of resolving such reference is known as reference resolution. In the above example, "*She*" and "*Her*" referring to the entity "*Genet*" and "*the institute of engineering*" referring to the entity "*AAiT*". These are two examples of reference resolution. To do this task, there is one kind of reference resolution i.e. coreference resolution from NLP applications. This task is used to resolve pronouns, proper names and nominal with clustering these entities together that they referring to [2]. The entities resolved by this system may be a person, place, organization, or event. In coreference resolution we should know the antecedent mention that is referent in the text. In coreference resolution task referent is the object that is being referred to. For example, "Genet" is the referent in the above example. An other term in coreference resolution is referring expression [30].

Referring expressions are the mentions or linguistic expressions given in the sentence and two or more referring expressions that refer to the same discourse entity are corefer. To extract referring expressions, referent, corefering expressions and entity look at the given example.

Example: *Mark Zuckerberg was born in May 14, 1984. He is the founder of Facebook. The 37 year old is widely known as the co-founder of Meta platforms.*

Referring expressions: *Mark Zuckerberg, He and the 37 year old*

Referent: *Mark Zuckerberg*

Corefering expressions: {*Mark Zuckerberg, He*}, {*Mark Zuckerberg, The 37 year old*}

Entity: *Mark Zuckerberg, May 14, 1984, Facebook, and Meta*

Therefore, coreference resolution is one of the most essential Natural Language Processing (NLP) tasks. The task of coreference is so complex to human and computer. Especially the task of coreference resolution is very challenging for computers when we compared to humans. Even though, the task of coreference resolution hard, it can be applied to various NLP tasks such as information extraction, information retrieval, question answering, machine translation and text summarization [34, 35, 36, 37].

As in many areas of NLP, the early methods for coreference resolution were built with machine learning approach [10], hand-crafted rules, transformer based approach (BERT) and deep learning approach. In the past decade, machine learning methods have shown advantages over knowledge-based ones and BERT model is one of state-of-the-art model for coreference task in the early period of 2019 [16]. Coreference resolution is not done for Amharic text using BERT. Let us take an example for English language to verdict mention pairs using bidirectional encoder representation from transformer. From Example 1 below "*Barack Obama*", "*his*", and "h*e*" are coreferents and from Example 2 "*Hillary Rodham Clinton*", "s*ecretary of state*", "*her*", "*she*", and "f*irst* l*ady*" are coreferents in the text.

Example 1: *Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.*

Coreferential: *{Barack Obama, his, He}*

Example 2: *Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.*

Coreferential: *{Hillary Rodham Clinton, Secretary of state, her, she, First Lady}*

There are three types of nouns phrases chosen for referring to an entity. These are proper names, nominals and pronouns.
Proper names are names of specific entities. For instance, “Bahir Dar” and “Ethiopia” are proper names in the given Example below.

Example: *Bahir Dar is one of my favourite places to visit. This city is found in Ethiopia.*

Nominals are noun phrases that have a noun as their head. Nominals cannot stand alone as a sentence but it contains a verb. Nominals are also called common nouns. For instance, from Example 1 “*a nice cup of coffee*” is nominal whereas “*a wall*” and “*the headquarters*” are also nominal from Example 2.

Example 1: *We all drink a nice cup of coffee.*

Example 2: *We found a picture on a wall in front of the headquarters.*

Pronouns is a word that can be replace noun in the sentence. It can be from one of the following categories: reflexive, definite and demonstrative.

Reflexive pronouns are one part of pronouns that refers back to a person or thing. Words like myself, yourself, himself, herself, itself, ourselves, yourselves and themselves are reflexive pronouns. For example, *John couldn't help himself.* “*himself*” is reflexive pronoun.

Definite pronoun are the second category of pronoun that refers to a specific noun. Some of the examples of definite pronouns are “I”, “me”, “you”, “she”, “he”, “it”, “we”, “us”, “they”, and “them”. For example, *my father was John Smith. He was seventy years old.* Here “*He*” is definite pronoun.

Demonstrative pronouns are the third category of pronouns that point to specific objects. They used in place of noun, noun phrase, activity, or situation. Demonstrative pronouns consist of this, these, that, and those. For example, *I appreciate all that you have done.* So here “that” is demonstrative pronoun.

2.3.1 Referential Expressions in Coreference Resolution

In discourse coreference chain or cluster is a set of corefering expressions. One task of coreference resolution is deals about linguistic properties characteristics of the language [44]. Knowing of these language characteristics of the coreference relation helps us to determine how to correct perform coreference resolution with minimizing the rate of errors. These properties of a language may vary from language to language because language has its own structure and rules. Some of the linguistic properties are number agreement, gender agreement, recency, grammatical role, pronoun, noun, verb semantic and so on [32, 38].

Number agreement is the referencing expressions in the discourse that agree in number. Example of number agreement is given below.

Example: *Barack Hussein Obama was born in 1961. He was the first African-American president of the United States.*

Here, "*Barack Hussein Obama*" is singular. Hence, the referent used is "*He*" in order to refer to the entity "*Barack Hussein Obama*". We use *"He"* to reference *"Barack Hussein Obama"*. Thus, *"Barack Hussein Obama"* and *"He"* agree in number.

Gender agreement is implies that the referencing expressions should be agreed in gender. Let us take an example to demonstrate gender agreement in a sentence.

Example: *Abebe has a plane. It is so expensive.* Here, "*plane*" and "*It*" agree in gender.

In grammatical role we can determine the hierarchy of candidate entities based on their grammatical role. This property takes advantage of the inherent grammatical nature of a sentence which gives more saliency value to subject entity as compared to an object entity. In other words, we assume that an entity which is a subject is usually more important than an object entity.

Example: *Abebe went to the Toyota dealership with Kebede. He bought a Corolla.*

In this sentence, we have "*Abebe*" and "*Kebede*" as candidate referents for the word "*He*". Here "*Abebe*" is the subject while "*Kebede*" is the object. Thus, keeping in mind the saliency, we deem "*Abebe*" to be coreferent to "*He*" rather than "*Kebede*".

In recency entities introduced in recent utterances are more salient than those introduced further back. Example: *Abebe has a Toyota car. Kebede has a BMW. Aster likes to drive it.* Here "BMW" and "It" are are coreferent based on recency.

Pronoun is a word that can replace a noun in a sentence and used instead of a noun or noun phrase in a sentence. The noun that is replaced by a pronoun is called an antecedent. For example, in the sentence *Abebe went to Entoto park, and parked next to a Toyota car. He went inside and talked to Kebede for* *more than an hour,* the word "*He*" is a pronoun that replaces the noun "*Abebe*".

Noun is a word that refers to a person (Obama), place (Addis Ababa), animal (dog) or things (table*)* in a sentence. Noun is not a pronoun. A noun phrase is a collection of words, usually a noun in addition to a modifier. For , in the sentence *that new blue pen is mine*, the phrase "that new pink blue" is the noun phrase. "Pen" is the noun, and the other words describe the Pen. Noun phrase can be classified into two: indefinite noun phrase and definite noun phrase.

An indefinite noun phrase is a noun phrase that is an indefinite description phrase. The indefinite article (a, an) is used before a noun that is general or when its identity is not known.

Example: *I saw a Toyota car today. Some Toyota cars were being unloaded. I saw this awesome Toyota car today.* Here *"Toyota car", "Toyota cars" and "Toyota car"* are coreferent.

A definite noun phrase is a determined noun phrase whose head is a noun with definiteness. The definite article (the) is used before a noun to indicate that the identity of the noun is known to the reader.

Example: *I saw a Toyota car today. The car was white and needed to be washed.* Here *"The car"* in the sentence is a definite noun phrase. "*Toyota car*" and *"The car"* are coreferent.
The last linguistic property is verb semantics. In the discourse some verbs tend to adapt more meaning to one of their arguments as compared to others while performing semantic analysis.

Example 1: *Abebe assisted Kebed. He was the designer of the project.* Here, the usage of verb "*assisted*" implies that the probability of "*Abebe*" being the designer of the project is higher than "*Kebed*". Thus, "*He*" refers to "*Abebe*" in the sentence.

Example 2**:** *Abebe condemned Kebed. He was the architect behind the project.* In this sentence, the usage of verb "*condemned*" implies that the probability of "*Kebede*" being the architect behind the project is higher than "*Abebe*". Thus, "*He*" refers to "*Kebede*" in the second sentence.

2.3.2 Coreference Link Type

Coreference resolution involves two main subtasks such as finding referring expressions in the text and clustering them into coreference chains. There are two types of coreference chains for coreference resolution task. These are: Identical (IDENT) and Appositive (APPOS).

Identical (IDENT)

This is the type of coreference link that contians names, nominal mentions, and pronominal mentions of the same entity. There is no restriction on what semantic types of noun phrase entities can be considered for coreference. For example:

As you know, <COREF ID="1" TYPE="IDENT">I </COREF> usually provide the scientific and technological entertainment in <COREF ID="2" TYPE="IDENT">our </COREF> meetings but on this occasion, <COREF ID="2" TYPE="IDENT"> our </COREF> Chairman suggested that <COREF ID="1" TYPE="IDENT"> I </COREF> present <COREF ID="1" TYPE="IDENT"> my </COREF> own personal view on events in <COREF ID="2" TYPE="IDENT"> the part of the world from which <COREF ID="1" TYPE="IDENT"> I </COREF> come </COREF> .

Appositives (APPOS)

An appositive is a type of corefference link that contains a noun phrase which modifies an immediately-adjacent noun phrase. The subsequent noun phrase separated only by a comma, colon, dash, or parenthesis. In appositives the noun phrase points to a specific object/concept in the world.

For example:

<COREF ID="1" TYPE="IDENT"> Barack Obama </COREF>, <COREF ID="1" TYPE="APPOS"> the former American president</COREF> has told <COREF ID="2" TYPE="IDENT"> the country </COREF> <COREF ID="1" TYPE="IDENT"> he </COREF> is ready for a long vacation.

2.3.3 Coreference Resolution Models

Researches performed with various models in the area of coreference resolution to get more effective result in time to time. There are many kinds of model for coreference resolution task. Some of them are mention-pair models, mention-ranking models and entity-based models. We briefly describe each of the models in the following sections [29].

Mention-Pair Model

This model is the most common coreference model since [29, 56]. The main feature of this mention-pair model is analysing individual properties of mention and its antecedent as well as the feature analyse the pairwise relation between mention. The given pair of mentions is classified as either coreferent or non-coreferent in mention-pair models. In this model, mention-pairs are made by taking into account each mentions and candidate antecedents. So we consider each mention as anaphor and the previous mentions as candidate antecedents.

Example: *Abebe lives in Ethiopia. He is very happy in that country.* Thus, mention *"Abebe"* is the best candidate antecedents of *"He"* and mention *"Ethiopia"* is the previous mentions of *"that country"*. So from this it is possible to make coreference chains with the help of clustering algorithm and pairwise decisions such as *{Abebe, He}*, *{Ethiopia, that country}*. The clustering algorithm can be done by concatenating each mention to its closest previous antecedent that is labeled as coreferent [5]. The clustering algorithm also works by connecting each mention to its highest scoring previous antecedent with a coreferent label [5]. In addition to this it can construct coreference chain by merging all mention-pairs that are labeled as coreferent [56].

Thus, suchlike clustering methods only focus on compatibility of each individual mention-pair. Therefore, with this method dependencies beyond mention-pairs will be eliminated. For example, a pairwise model may detect both (Ms. Matewal, Mastewal) and (Mastewal, He) pairs as coreferent. However, the model could have inferred from the first pair that “Mastewal” is a female name and therefore it cannot be coreferent with “He”. Accordingly, to overcome this problem in mention-pair model, other clustering algorithm is developed. From this the first one is Bell-tree clustering; the second one is graph partitioning. Bell-tree clustering method works by creating entities from individual mentions as a tree. The root of a tree is an entity that holds the first mention of the sentence. One mention should be processed at a time during the creation of each tree level. Each candidate anaphor is paired with all the preceding candidate antecedents or starts a new entity. So, the second mention can be linked to the first previous mention at the root node or starts a new entity. A leaf of the tree is represented by the possible outcomes of coreference resolution. This approach addresses the clustering step of a coreference resolution process as its input. It assumes a set of mention pairs for a given document, labelled as positive (two mentions co-refer) or negative (two mentions do not co-refer) by an external classifier [8]. Therefore, we do potential links between any of the mention as positive.

Therefore, the probability of linking a mention ’m’ to a partial entity ‘e’ is estimated by this equation.

$$\Pr(\text{link}|e, m) \tag{2.1}$$

The usage of equation 2.1 is for scoring each decision while we apply it on the output of a mention entity model. The second is graph partitioning algorithm that creates a graph which contains node and edges. Each node is mention and edges are score that is calculated based on the coreference probability of the corresponding nodes. Various graph clustering algorithms is used to find partitions of mentions.

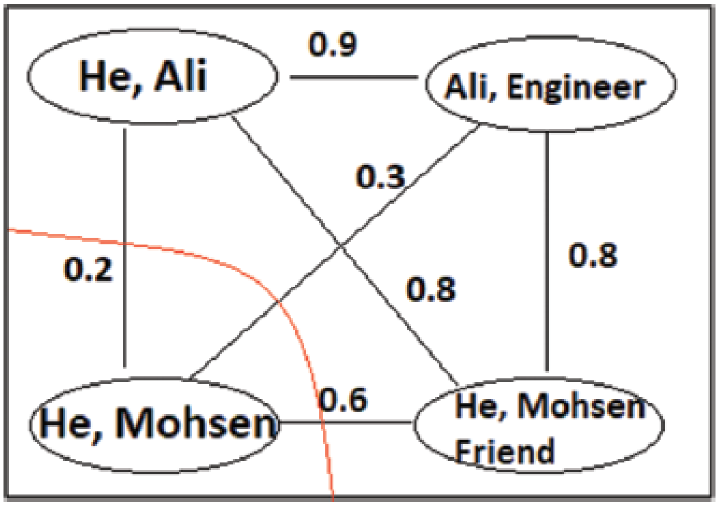


Figure 2.1: Coreference graph

Mention-Ranking Model

Mention ranking model is one of the most useful approaches in coreference resolution task to get effective result with effective datasets. This way, the model is the most powerful and popular coreference models in the task of coreference resolver system right now.

The neural mention ranking model of a deep learning algorithm takes inputs to process it using coreference system that produce a desired result. These inputs are mention and candidate antecedent in which it is any mentions that can be present before a mention in the sentence [55].

Thus, the model produces a score for the pair of mention and candidate antecedent. In this model the mention may have correct antecedent which is coreferent or may not have intended antecedents which are not coreferent. The coreference system can cluster the mention and the candidate antecedent that produces a score when the chain is correctly formed.

The model has various layers to score a pair of mention and antecedents in the document. These layers are hidden layer, scoring layer and input layer. In input layer the neural network takes the extracted mention head words and all words in the mention for each mention in the text. Thus, each word is represented by a real value vector.

The main aim of mention ranking model is to increase the mention pair model of coreference relations by taking into account all antecedents of a single mention. There is a comparison among various candidate antecedents of a single mention.

This kind of model is so fast and simple to train the model with reliable dataset. In addition to this there is no need to determine informative mention-pairs manually. The informative pairs can be automatically identified during the learning process. Mention-ranking models score pairs of mentions for their likelihood of coreference rather than comparing partial coreference clusters. For example consider this example. As you can see from the example gold mentions are enclosed in square brackets. Mentions with the same text are marked with different indices and the indices in parentheses denote to which key entity the mentions belong.

Example 1: *Barack Obama, the former US president, has told the country he is ready for a long vacation.*

Example 2: *These items, which were the pride of the Ocean Park, have made this place the most popular tourist attraction in Hong Kong for some time. However, since Disney entered Hong Kong, the Ocean Park, sharing the same city as Disney, has felt the pressure of competition.*

A mention-pair model tries to learn all pairwise relations including the underlined word "*this place*", "*the Ocean Park*". However, these pairs are not informative pairs to learn from. On the other hand, a ranking model does not enforce the classifier to learn from all coreferring pairs. Instead, the model can only concentrate on learning from more informative pairs, for example "*the Ocean Park*", "*the Ocean Park"* in the occurrence of the mentions from the beginning to end in the Example 2 respectively.

Yang *et al.* (2003) proposed a simplified ranking model by considering all candidate antecedents together. The model compares pairs of candidate antecedents at a time to identify which of them is a potential antecedent for a given anaphor. If the candidate antecedent wins most of the pairwise comparisons, we will select it as the antecedent of an anaphor.

Different methods are proposed for training most effective ranking model, these are during training of the model, select the best antecedents through learning process heuristically and Learning from best antecedents during training, selecting best antecedents automatically.

- During training of the model, select the best antecedents through learning process heuristically. In this process, the model learned with various parameters to identify the closest true antecedent of a mention based on high score as compared to other antecedents.
- Learning from best antecedents during training, selecting best antecedents automatically. The mention-ranking models are used by Chang *et al.* (2012), Wiseman *et al.* (2015), and Clark & Manning (2016a) are examples of such methods. At the time of this model, the model selects true antecedent based on the highest-scoring that is $\hat{t} = \arg\max_{t \in T(m_i)} \text{score}(t|m_i)$. The scoring methods pick the best antecedent during learning. The antecedent selection of all mention conducted after we connect each mention to its highest scoring antecedent. This model is known as antecedent-trees because the result produces a tree structure of a candidate antecedent. Thus, from the tree the parent of each node is the selected antecedent from all candidates.

Generally, the mentions are selected by allocating each mention first with its uppermost scoring candidate antecedent.

Entity-Based Model

Entity-based models are proposed by Luo *et al.* (2004), Yang *et al.* (2008), and Wiseman *et al.* (2016) and so on, that learns each mention of the text in a left-to-right method to merge each of the identified antecedent mentions to the potential emerged entity. But an entity-centric model merges two partially identified entities in the discourse. The most advantage of this model over other models is that it includes more advanced features that are defined over entities instead of mentions or mention-pairs [58].

The entity-based models define based on entity-based features. Entity based features can be made by applying all and most coarse quantifier predicates to mention based features. As an example, from mention type noun, we can create all-nouns are true which indicates that all mentions of the examined cluster are nouns. This features can also be create by combining properties of individual mentions [16].

For instance, proper name-pronoun is an entity-based feature which is constructed based on the property of mention type with clustering one proper name and one pronoun. As Wiseman *et al.* (2016) proposed this model by applying an LSTM network on entities in which entity based features are trained automatically and implicitly. Generally, this approach is the most complex and practically the least successful model in coreference resolution task. Research works shows that the results of the entity based model are slightly better than those of their mention ranking model.

2.3.4 Coreference Resolution Techniques

In coreference resolution task there are four kinds of approaches. These are rule-based approach, machine learning-based approach, transformer based approach and deep learning-based approach. For this task those approaches has its own strength and weakness. In the following section, we addressed those approaches to determine their strengths and weaknesses for coreference resolution task.

Rule-Based Approach

The rule based approach is a type of technique that uses set of rules which prepared by humans manually. Since these sets of rules are often based on pattern-matching, rules are also thought of as extraction patterns. To be able to write good extractors, humans want to know what types of features are involved in a text. If rule-sets is complex, it will be less readable to humans. Therefore this leads to more costly to manage, debug, and extend the approach. Because of this, rule-based coreference resolution approaches have well-known difficulties in scalability. This method is so hard to generalize it to other domains because new hand-crafted patterns would be needed. Thus, to solve these problem machine learning approach was invented for coreference resolution task.

Machine Learning-Based Approaches

Machine learning approaches uses large amount of dataset to get high performance of the system during training phase. It classifies using a large set of positive and negative training examples.

The main challenges in machine-based coreference resolution is it requires many feature-set during the process of experimentation. Many feature must be identified before building the model. The other challenge in machine learning approach is large amounts of data need to be prepared for training and evaluation. In the following, we introduce some approaches to machine learning-based coreference resolution such as supervised, unsupervised, and semi-supervised, which address this problem in different ways [45].

Supervised machine-learning approach classifiers are trained using fully hand-labeled data where entities and the link between them were annotated. This involves labeling data in a corpus with entities and relationships between them. This approach combines various lexical, syntactic, and semantic features with machine learning classifiers. This method is most important for automatically evaluating different feature sets and classification approaches using cross validation. Even though these approaches can achieve high performance, they need expensive data annotation and the selection of features. The other disadvantages are the high costs associated with manually labeling gold data if none exists. Generally, supervised learning for coreference resolution is preferred over unsupervised learning due to ease of evaluation [54].

Unsupervised approaches use massive amounts of data and extract more mentions, and the subsequent associations may not be easy to map to the mentions required for a particular knowledge base.

Unsupervised approaches can extract mention. The main drawback of this method is that it is not possible to predict what mention will be found, and human judgement is needed to determine whether a found mention is informative. The system is not very useful for a specific kind of mentions, and it is often not possible for the system to label the mentions in a way that is informative for a human. Another limitation of the unsupervised approaches is that the mentions they produce might not be compatible with mentions that exist in knowledge bases.

Semi-supervised approach is the third type of machine learning approach. The basic idea of semi-supervised machine learning approach is to label the unlabeled sample in the dataset, which mainly solves the problem of learning. It improves the ability of model generalization in the label sample. It can effectively reduce the dependence of human participation and corpus annotation. It has been widely used for large-scale text coreference resolution tasks.

Deep Learning-Based Approach

Deep learning approaches underlined the context of words accurately. In coreference resolution non-deep learning approaches are depends on NLP tools for feature extraction. This deep learning algorithms has the ability able to discover complex structures from large datasets using the back propagation algorithm that updates their internal parameters. This peculiarity of deep learning lets the model extract mentions with better accuracy. Recently, many researchers have applied depth learning techniques to coreference resolution [43]. For example Transformers: Transformers are first formulated by Google in 2017. At the starting stage of this approach, language models primarily used Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) to solve problems in NLP tasks. Although these models are very competent, the transformer is shows a significant improvement because it doesn't need sequences of data to be processed in any fixed order, whereas RNNs and CNNs do this thing. Transformers can process larger amount of data in any order. This, facilitated the invention of pre-trained models like Bidirectional Encoder Representation from Transformers (BERT), which was trained on massive amounts of language data prior to its release.

Transformer Based Language Understanding

Transformer is a deep learning model that has encoder-decoder network. Encoder part of a transformer architecture uses self-attention, whereas decoder side of a transformer architecture uses attention. Thus, the transformer architecture is an encoder-decoder network. An encoder reads the entire sequence of words at once without reading the text input sequentially (left-to-right or right-to-left) and a decoder produces a prediction for the task. This framework has simple architecture and it is so fast to train the model. The frameworks also understand complex patterns in the data with the help of attention mechanism. This is most accurate deep bidirectional model [16, 17, 23].

In general, the transformer itself is both the whole encoder and the whole decoder and they both consist of N-layers where each layer is the typical multi-head-attention layer with dropout and normalisation. In language models we often only use the encoder. The basic drawback of the previous works is inability to consider both left and right context of the masked word. To tackle this problem we use transformer based model, i.e. BERT which is able to achieve state-of-the-art performance on a variety of task [17].

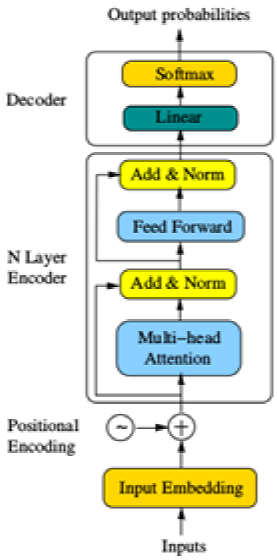


Figure 2.2: Transformer Architecture

Bidirectional Encoder Representation from Transformers (BERT)

BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers. It is the first NLP technique to rely solely on self-attention mechanism, which is made possible by the bidirectional Transformers at the center of BERT's design. It is also a neural network based technique for natural language processing task [57]. It was created by Google to help their search engine in understanding the content on various web pages. BERT model is very helpful to the various natural language processing application. So, it uses a special tokenization called WordPiece which can split words to sub-words [23].

When a word was split to several sub-words, only the embedding of the first sub-word was taken. This language model is a pre-training deep transformer model that uses encoder from transformer and it is trained based on a self-supervised approach [16, 17]. It makes use of transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text. As long as the transformer based model is non-direction/bidirectional because its self-attention layer performs self-attention on both right and left direction of the word. This property of a model is more important because sometimes words may change meaning as a sentence develops. The BERT uses a transformer encoder as shown in the figure 2.2. The transformer architecture consists of encoder (multi-head attention, feed forward neural network and normalization) and decoder (linear layer, softmax layer). However, BERT used the decoder of a transformer model to generate hidden value representation and transformer decoder (for generating text in NLP tasks) is substitute by a linear followed by softmax layer. The encoder consists of 12/24 layers. Those each layer has two sub-layers: multi-head self-attention mechanism and feed forward neural network. There is a connection around each of two sub-layers followed by normalisation. While we produce result/predict the next word in a given sequence of word in various model is very difficult. All the developed models cannot predict the masked word effective. To overcome this challenge, BERT uses two training strategies: MLM and NSP.

Masked Language Model (MLM)

Masked Language Model is reasonable to believe that a deep bidirectional model is strictly more powerful than either a left-to-right model or the shallow concatenation of a left-to right and a right-to-left model. Unfortunately, standard conditional language models can only be trained left-to-right or right-to-left, since bidirectional conditioning would allow each word to indirectly “see itself”, and the model could trivially predict the target word in a multi-layered context. The main aim of MLM is to hide a word in a sentence and predict what word has been masked/hidden based on the hidden word context. Therefore, in this model words in a text are randomly erased and substitute with a special token ‘masked’ with some small probability. The transformer is used to generate a prediction for the masked word based %on the unmasked words surrounding it, both to the left and right side. 15% of a words in each sequence are replaced with a ‘MASK’ token.

Next Sentence Prediction (NSP)

Initially during training the model, the model accepts pairs of sentences that is take as input. The model learns to produce result if the second sentence in the pair is the following sentence in the real document. At the time of training 50% of the inputs are a pair in which the second sentence is the subsequent sentence in the real/original document, while in the other 50% a random sentence from the corpus is chosen as the second sentence. The basic assumption of this idea is that the given random text will be separated logically from the first sentence. In order to overcome this problem BERT use NSP. So as in general the main aim of the NSP training process is that whether the given two sentences has a logical, sequential connection or not.

2.4 Evaluation of Coreference Resolution

Evaluation is the process of measuring the performance of the given system by various metrics. It is the tangible evidence of an experiment. To do that we used standard coreference resolution metrics score. MUC B^3, CEAF, BLANC and CoNLL are the major evaluation metrics for coreference resolution task. Those metrics are expressed in terms of precision, recall and F-measure. The metrics evaluated the outcomes of the coreference system based on Key and response entities. Key/gold entities are expected entities, response entities are those entities that is generated entities from coreference system. MUC metrics is link based entity between two mention. B^3 metrics measures the mention in coreference resolution task. CEAF metrics measures the system by mapping one key entity to one response entity, because there may be repeated mention in B^3 metrics. The CEAF metric overcome this problem. The other evaluation metric is BLANC, it measures the coreference link and non-coreference link. The last metric is CoNLL that measures they system by taking the average of the four metrics [18, 36].

Precision (P)

Precision is the ratio of correctly identified mentions (TP) to the sum of correctly identified (TP) and incorrectly identified (FP) coreference noun phrases i.e

$$P=\frac{TP}{TP+FP} \tag{2.2}$$

Recall (R)

Recall is the ratio of the correctly identified coreference chain (TP) to the sum of total correctly identified (TP) and unidentified coreference chain (FN) i.e.

$$R=\frac{TP}{TP+FN} \tag{2.3}$$

F- measure (F1)

F-measure is the weighted harmonic mean of precision and recall i.e.

$$F1=\frac{2*R*P}{P+R} \tag{2.4}$$

CHAPTER THREE: RELATED WORK

3.1 Introduction

In this chapter, we discusses some of the related research works which are done so far on coreference resolution. Among them, we have selected the most pertinent research for different languages, which are related to our work. In a long years ago coreference research is done using rule-based, machine learning [52], hybrid approach [39]. In recent years researchers proposed coreference resolution for English and other languages using BERT model. This work also uses this technique to obtain prominent experiment result. In this chapter, we present coreference resolution for Ethiopian and other languages using various approach.

3.2 Coreference Resolution Using Rule Based Approach

Majdi Beseiso and Abdulkareem Al-Alwani [13] developed a coreference resolution for Arabic using morphological features. In their experiment, both rule-based and machine learning approaches showed limitations in Arabic conference resolution. Because, rule-based approach requires a large set of rules and machine learning approach needs annotated corpus to be used as training and testing data. From their conclusion, machine learning can be used for improving the process of morphological analysis by learning new rules for the process. The experiment result shows that the precision value is 0.855, recall value is 0.886 and F-measure value is 0.87.

Alexander Wallin and Pierre Nugues [21] proposed coreference resolution for Swedish and German using distant supervision. The work is focused on Swedish and German language sentences. The researchers used distance supervision approach for Swedish and German text without manually annotated text. It is an end-to-end coreference resolver for Swedish and German languages. Feature sets are easily extracted from the Swedish Treebank [39] and the German Tiger corpus. They generate labelled training set using parallel corpora, English-Swedish and English-German, where they solved the coreference for English using CoreNLP and transfer it to Swedish and German using word alignments. To carry this out, they identify mentions from dependency graphs in both target languages using hand-written rules. Supervised learning [54] uses labelled data, often obtained through a manual annotation, whereas in the case of distant supervision, the annotation is automatically generated from another source than the training data itself. They used various metrics to evaluate coreference resolution. The three metrics independently evaluate the performance of the system such as MUC, B^3, and CEAF-e.

In Swedish the mentions are correspond to noun expressions. The identification of noun expressions in German proved more complicated than Swedish, especially due to the split antecedents linked by a coordinating conjunction. In Swedish, the rule for determining the same type of split antecedents only needs to check whether a conjunction has children that were noun expressions, whereas in the case of German language, German rule needed more analysis. Eventually, to estimate performance of the system the experimenter used training and testing datasets. They estimated the system using standard coreference evaluation score from the CoNLL-2012 shared task [19]. Finally, they attained a score value of MUC is 46.24%, B^3 is 28.87%, CEAFE is 27.68%, CEAFM is 32.21%, BLANC is 29.41% and CoNLL is 34.28% for Swedish language. The experiment result of German language is presented in the table below.

Language	Models	MUC	B^3	$CEAF_E$	$CEAF_M$	BLANC	CoNLL
German	J48	34.29	2.63	2.55	12.81	4.67	13.36
	Random forest	33.51	2.54	2.4	11.82	5.46	12.81
	Linear regression	33.97	2.36	1.35	12.3	4.58	12.56

3.3 Anaphora Resolution Using Knowledge-Poor Approach

Temesgen Dawit and Yaregal Assabie [11] have done research on Amharic anaphora resolution using knowledge-poor approach. They conduct research on Amharic text. During their implementation, they have been used pronominal anaphoric terms in anaphora resolution system. They identified hidden anaphors, independent anaphors and candidate antecedents. To evaluate the performance of the system, they collect Amharic text corpora from different sources (Amharic grammar book). They used 311 sentences to test the performance of the system. From these sentences, they identified a total of 315 verbs which contain hidden personal pronouns. In this sentence the number of independent anaphors is minimal, so they collected an additional text from the Amharic Bible where they selected a total of 163 sentences having a total of 110 independent personal pronouns. The performance of the system is evaluated using 10-fold cross validation technique. After they tested the Amharic anaphora resolution system, they obtained experiment result [11]. Based on the collected dataset, they achieved a success rate of 81.79% for resolution of hidden anaphors whereas an accuracy of 70.91% is obtained for resolution of independent anaphors.

3.4 Coreference Resolution Using Machine Learning Approach

Veena G *et al.* [4] proposed coreference resolution for English using Support Vector Machine (SVM). They used statistical approach for coreference resolution of English noun phrases in a document using SVM as classifier. The task is performed with two basic phases training and testing phase. The training phase includes preprocessing, feature extraction and learning algorithm tasks. The testing phase also has these tasks with coreference result. They drafted 10 features to frame the coreference resolution system. In this system SVM classifier is used to resolve whether two noun phrases are coreferential or not coreferential. The machine learning technique i.e SVM classifier performed better outcome when they compared to other machine learning models. The experiment achieved 84.7% precision value, 86% recall and 84.8% F-measure value.

Soon *et al.* [5] proposed a machine learning approach to coreference resolution of English noun phrases. They used various features to extract mentions in a text. Those features are build using rule based system. From these feature some of them are distance feature, i-pronoun feature, j-pronoun feature, string match feature, definite noun phrase feature, demonstrative noun phrase feature, number agreement feature, gender agreement feature and so on. They develop the system based on decision tree. The learning algorithm used in their coreference experiment is C5.

They evaluated the approach on common data sets, namely, the MUC-6 and MUC-7 coreference corpora [10, 18, 35, 47].

3.5 Coreference Resolution Using Transformer Based Model

Trieu *et al.* [23] proposed coreference resolution system using Bidirectional Encoder Representations from Transformer (BERT) and syntax-based. An English language is considered for this work. The task of this work is firstly, they filter noisy mentions based on parse trees with increasing the number of antecedent candidates. Secondly, instead of relying on the LSTMs, they integrate the highly expressive language model that is BERT into their model. BERT model is a contextualized language model that can efficiently capture context in a wide range of NLP tasks. In their work they consider the number of antecedents and filtering noisy to improve the performance of coreference resolution system. On both mention detection and coreference detection BERT filter achieved the best performance in all metrics. By using mention filtering they improved the baseline LSTM model from 4-16% points in F-score varied by metrics. Lastly, for coreference resolution system, they have obtained the best performing F-scores of 44%, 48%, 39%, 49%, 40%, and 57% on the test set with B^3 , BLANC, CEAF-e, CEAF-m, LEA, and MUC metrics, respectively.

Arthi Suresb *et al.* [15] developed coreference resolution for English language using BERT. pronoun coreference resolution with the help of BERT. They develop a model with the help of English language dataset. They explore various models for coreference resolution task. Among this model transformer model one of the state-of-the-art models for English corefernce resolution. BERT transformer model is used in a variety of architectures for the coreference resolution task. First it is used as an input for a rule-based heuristic, and second used as embeddings in a mention-pair ranking architecture, third used as a replacement for a long short-term memory network in an end-to-end neural model [32] that jointly learns mention extraction and coreference clustering as well. They hypothesize that BERT (bidirectional encoder representations from transformers), learned with the objective of predicting the original vocabulary ID of a randomly masked word based only on its context. In this system they used softmax of self-attention matrix to know the correct antecedent of a target mention by calculating the weight of each mention. The performance of the system is evaluated using standard metrics and algorithms. The mention ranking algorithm using BERT achieves an overall F1 of 76.0.

Wilkens *et al.* [34] proposed coreference for spoken and written French language by the help of end to end system as a baseline. They developed a coreference resolution model that outperforms the current state of the art for French. It is the first French coreference resolution model trained on written text. To evaluate the coreference resolution approaches they used both written and spoken French corpora.

To build this system effectively they studied three coreference approaches: the mention-pair, easy-first and neural network approaches [57]. As they said, all of them are very important for French language. The mention pair model uses binary classifiers to find pair of mentions that are coreferent/not. The second easy-first approach is the rule based system that uses various sieves to detect specific relation types.

They trained the model and test the performance of the coreference system using two corpora (spoken language and written texts) that are annotated with coreference chains, and augmented with syntactic and semantic information. They obtained 84.13%, 80.09%, 87.91%, 85.04%, 78.11% and 79.28% MUC, B^3, CAEF-e, CoNLL, BLANCand CAEF-m experiment result respectively.

3.6 Coreference Resolution Using Deep Learning Approach

Lee *et al.* [32] explored end-to-end coreference resolution system for English. They develop the system based on deep learning approach that is Bidirectional Long Short Term Memory (BiLST). English dataset is used from the CoNLL-2012 shared task for training and testing the system [19]. This dataset contains 2802 training documents, 343 development documents, and 348 test documents. The training documents contain on average 454 words and a maximum of 4009 words. The performance of the system is measured in MUC, B^3 and $CEAF_{\varphi 4}$. The value of F1 for each metrics are 77.2%, 66.6% and 62.6% respectively. The average F-measure value is 68.8%.

Kupriianova *et al.* [55] developed the mention ranking approach to corefrence resolution for Russian language. As they discussed on their paper the task of coreference resolution is to find and group all the mentions in the text according to their referents. Mentions are typically represented by noun phrases (NPs), named entities and pronouns in this task [57]. Similar mentions are paired together to make coreference chain. In the pair of mentions the first one is the antecedent, whereas the second one is an anaphor. Their system presents about grouping a mentions into clusters, which refer to a single real-world entity. The system is developed based on neural networks mention-pair and mention-ranking model [51]. Mention-pair encoders transform a pair of mentions and its antecedent into their distributed representations. The mention-pair encoder contains a mention ¨m¨ and an antecedent ¨a¨ that is developed as a feed forward neural network with three hidden layers of rectified linear units (ReLU). These three hidden layers are hidden layer h_1 hidden layer h_2 and hidden layer h_3 scoring Layer. The final layer is a fully connected layer of size 1:

$$h_i(a, m) = \max(0, W_i\, h_{i-1}\,(a, m) + b_i\,) \tag{3.1}$$

Where h_i (a, m) is an output of the *i*-th hidden layer for a pair of mention ¨m¨ and its potential antecedent ¨a¨. W_i is a weight matrix and b_i is the bias for the *i*-th hidden layer. The input layer of the mention-pair encoder takes a vector of features of a mention and its potential antecedent as well as additional pair features. The output of the last hidden layer is the distributed representation of the pair which is used as an input to the mention-ranking model.

The purpose of the mention-ranking model is to estimate the score of coreference compatibility for the pair of a mention ¨m¨ and its potential antecedent ¨a¨. To compute this score one applies one fully-connected layer to the distributed representation of the pair r_m (a, m),

$$s_m(a, m) = W_m\, r_m\,(a, m) + b_m \tag{3.2}$$

Where $s_m(a, m)$ denotes a score of coreference compatibility of a pair of mention ¨m¨ and its potential antecedent ¨a¨, $r_m(a, m)$ is the distributed representation of this pair.

In their work, they also presented an approach using neural networks based on an adapted version of the mention-ranking model [57]. They evaluated the system by improving the state-of-the-art F1 score from 0.63 to 0.71, measured with the B^3 metric. The experiment result shows that 71.70% of precision, 70.92% of recall and 71.31% of F1 score.

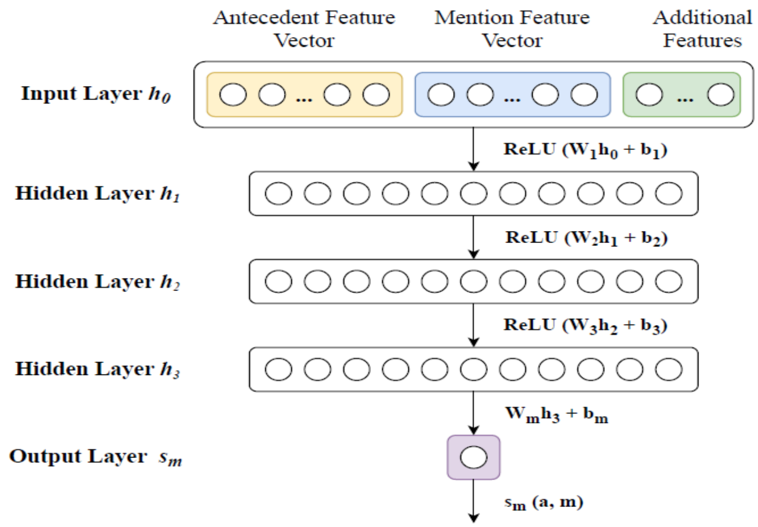


Figure 2.3: Neural Network Topology

3.7 Summary

In this chapter, we reviewed various study attempts to develop coreference resolution sytem for European and Asian languages as well as Ethiopian language with their explicit methods that they used, and evaluation outcome that they obtained from their experiment. The review also presented that various approaches are applied for coreference resolution task such as machine learning approaches, deep learning approaches, transformer based approaches and so on. Coreference resolution has the potential of employing transformer learning models with the creation of huge datasets. Among the various transformer language model type's bidirectional encoder representation from transformer (BERT) is most common in coreference resolution research works. As we discussed above, BERT is adapted for coreference resolution task in different languages effectively with achieving high performance of the model. The coreference resolution researches are language dependent. Therefore, various related works have been reviewed. From the related works we see that language specific and each language require unique technique and algorithm. Thus, we hypothesis that transformer encoder which is BERT language model can achieve high performance for Amharic coreference resolution model. Accordingly, in this study we use BERT to develop coreference resolution for Amharic text.

CHAPTER FOUR: THE PROPOSED SOLUTION

4.1 Introduction

As we can see the previous chapter, different approaches have been discussed to resolve coreference resolution for various languages. All the developed models are depending on the characteristics of the languages. We proposed coreference resolution for Amharic text using BERT model. Aim of this proposed model is to detect mentions by automatically learning features from input data. This chapter presents about system architecture with detail description. The first part of this chapter discusses about architecture of Amharic coreference resolution model. The next part of this chapter is text pre-processing steps such as cleaning, tokenization and sentence segmentation. Lastly, the testing phase and coreference resolution are discussed.

4.2 System Architecture

In this section, we designed the architecture of the Amharic coreference resolution model as shown in Figure 4.1. The architecture shows the flow of components and subcomponents starting from input to output of the Amharic coreference resolution process. In this study we proposed a transformer-based approach called Bidirectional Encoder Representation from Transformer (BERT) model. In this model word embedding and BERT are used to better capture local features. The proposed model contains different components. These are the pre-processing, word embedding, feature extraction and coreference resolution. These components are briefly described in the next section.

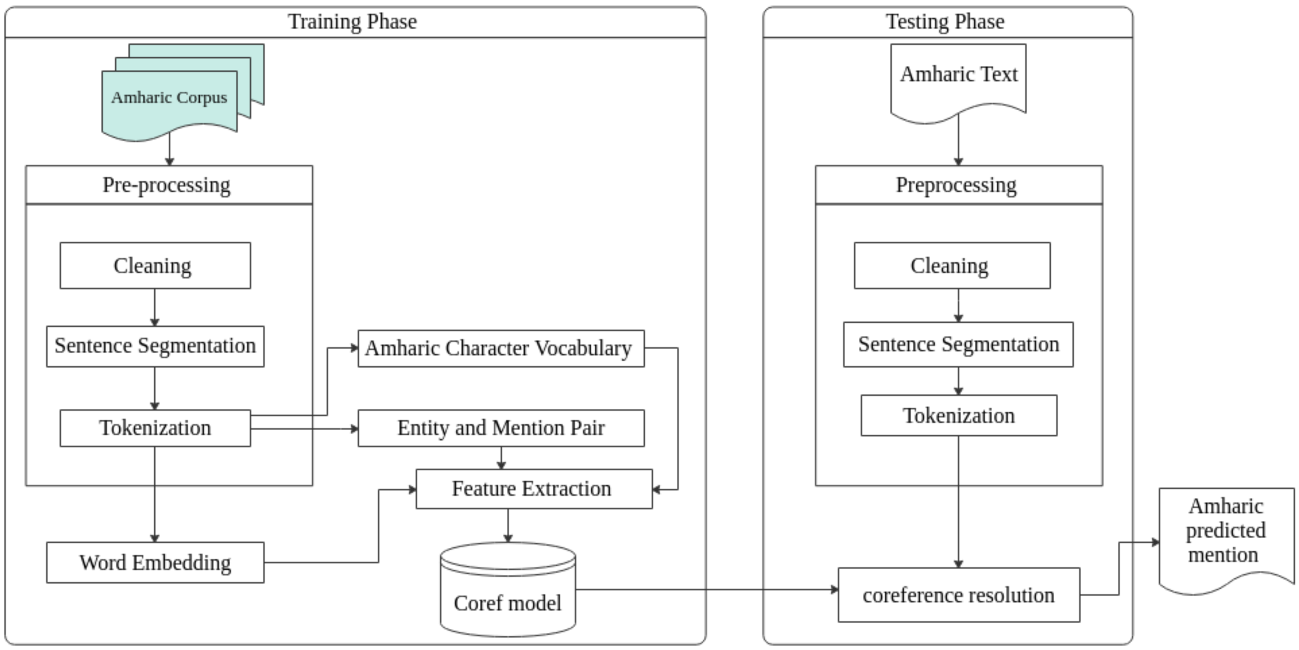


Figure 4.1: Amharic Coreference Resolution Model

4.3 Training Phase

4.3.1 Preprocessing

Accessing of prepared corpus in Amharic language is very challenging in coreference resolution task. In this study, first we gather Amharic sentences from various reliable sources. These Amharic text data is preprocessed before we fed into the system. Pre-processing is used to clean data, which is make it suitable for further analysis, because all words are not necessary to build word embedding. Next to that we prepare a dataset in CoNLL format [19]. This corpus used as input for training and testing the system in coreference resolution task. From the total corpus 80% of the corpus are assigned as train data and 10% of the corpus are assigned as test data using train/test splitter method as well as 10% of the corpus are assigned for validation. Two kind of training datasets are used in this system: train data and validation data, those dataset are labeled. The other datasets are grouped under word embedding and labeled as testing data. In training phase the pre-processing component includes sentence segmentation, tokenization and cleaning of unnecessary numbering format. Generally, in this work we prepared coreference resolution corpus for Amharic language from scratch.

Cleaning

In this stage of the model we removed non Amharic characters and unnecessary punctuations from the prepared Amharic corpus, because the collected data may have various problems like misspelled words, incomplete words, and unformatted data. Unnecessary numbers also cleaned from the corpora, because it reduces the performance of the proposed model. We also checked the collected sentences if there is misspelled Amharic words. We inserted cleaned Amharic sentences to build word embedding. If we do pre-process highly, the performance of coreference resolution will be high. The cleaner works according to Algorithm 4. 1.

Algorithm 4.1: Cleaning of Amharic Text Algorithm

Input: Amharic_corpus

Output: Cleaned_Amharic_text

```
Read list of Amharic_character in Amharic_corpus
For J=1 to number_of_character in Amharic_corpus
   If (Amharic_corpus[J] ! == Amharic_character)
         Remove Amharic_corpus[J];
  End If
 End For
```

Sentence Segmentation

It is the process of dividing free-flowing text into meaningful sentences with plenty of advantage for any natural language processing application. Many times it is the big challenge in natural language processing of deciding where sentences begin and end. This contradiction may come from the potential ambiguity of punctuaition marks. The sentence splitter recognizes the sentence based on Amharic punctuation marks that are present at the end of the sentences. In written Amharic, four dots (::) indicate the end of a sentence. In addition to this, sometimes we may use exclamation mark (!) and question mark (?) to indicate the end of the sentence. The usage of this task for coreference resolution is that, it is possible to identify the distance of each mention between sentence i.e. distance feature. If ¨*a*¨ is the first mention of the first sentence and ¨*m*¨ is the second mention that is present in the second sentence. The feature captures the distance between ¨*a*¨ and ¨*m*¨. If ¨*a*¨ and ¨*m*¨ are in the same sentence, the value will be 0. If they are one sentence apart, the value is 1; and so on. The possible values are 0, 1, 2, 3, and so on. Algorithm of Amharic sentence segmentation is given below in Algorithm 4. 2.

Algorithm 4.2: Amharic Sentence Segmentation Algorithm

Input: Pure_cleaned_Amharic_text

Output: separated_Amharic_text

```
use Algorithm1
    For every sentence in Pure_cleaned_Amharic_text
    Find punctuation marks :: or ! or ?
        If punctuation marks found
            Separate sentences
        End if
    End for
```

Tokenization

The next and vital stage in text processing part is tokenization. It is the process of splitting a given sentences into separate word called tokens. Token is the smallest unit that will be extracted from the input sentence. Thus, tokenization takes the input text supplied from a user and tokenizes it into a sequence of tokens, which is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called token and finally it gives the tokens to the next phase. In Amharic language tokenization step is done by finding word boundary, which is the ending point of a word and the beginning of the next word in the sentences. Amharic tokens may be contains words, punctuations, mumbers and so on. In order to find the boundaries of the sentence for our language, we find white spaces and punctuation marks as boundary markers. The Amharic language has its own special punctuation marks which demarcate words and sentences in a stream of characters such as ‟colon‟ (:), ‟four dots‟ (::), ‟comma‟ (፣), ‟exclamation mark‟(!), ‟question mark‟(?)...etc. Therfore, ‟colon‟ (:) is used to detect the end of the word in Amharic language. Most of the time ‟question mark‟ (?) and ‟four dots‟ (: :) are used to detect the end of the sentence. Thus, in this study we used NLTK word splitter to split Amharic sentences into smallest unit based on Amharic punctuation marks and white spaces. The tokenizer works according to Algorithm 4.3. For instance, look at the given sentence below:

Example: *ጠቅላይ ሚኒስትር ዐቢይ ወደ ጦር ግንባር ዘመቱ።*

Algorithm 4.3 : Amharic Tokenization Algorithm

Input: cleaned_Amharic_text

Output: List of Amharic_words

Read list

detect punctuation symbols() and white space

If file in read ends with white space or :: or ! or ?

Get sentence

Split sentences into tokens

Put each word in tokens

End If

4.3.2 Word Embedding

The embedding method is done to obtain the word embedding vector for each mention pair and comparing the similarity of them. Embedding also used for examine the similarity of mention sentence embedding with the candidate entity.

Word embedding is the process of mapping each word to the correct context representation in the text [38]. Training of word2vec requires large amount of words. In Amharic language there is no trained word2vec embedding. In this study we processed the unlabelled text for the word embedding technique that is used to map each word with its context. The input data is huge amount of data, that we apply preprocessing step like tokenization to transform it into vectors by looking up word embedding before we used it for training a model. From the word embedding, we gain context of words in a text by the interconnection of the different conceptual terms. In this vector representation stage semantic and syntactic word relation are easily learned. The word-embedding developed based on Algorithm 4.4.

Algorithm 4.4: Training Word embedding algorithm

Input: Amharic dataset

Output: Amharic word embedding model

use Algorithm1, 2 and 3

Read Amharic text into a list of strings

Tokenize the Amharic_text ()

Word Vector Function()

Build()

Save the trained Model

4.3.3 Feature Extraction

Various learning algorithm is adapt in NLP to obtain an in-depth analysis of the sentence. To analyse the given sentence feature must be extracted. Feature extraction is the process of determining and extracting all the necessary features from the input data by feature extractor. In each document, features extraction is performed after we finish the preprocessing stage and mentions detection.

To extract features first we recognize the entities, and then mention pairs are identified to extract necessary features from them. Therefore, important features are taken from mention pairs. In the feature extraction process we use Amharic character vocabulary, word embedding and training data. Features are extracted before training or evaluating the model, this step said to be “bertification”, since we use Bidirectional Encoder Representations from Transformers (BERT). The network of BERT can identify context of a token from both the right and left side of the word at first layer and go through to the last layer. While using BERT (deep transformer based model), context of a word is represent to sub-words in each words. The feature extractor use distance and exactly string match features.

Distance feature captures the distance between coreferent terms. If mention “m” and its antecedent “a” present in same sentence, the value is 0. If it exist in one sentence apart the value will be 1.

Exact string match feature links mentions with exactly the same extent word/phrase in the text by considering them as coreferent. If two mention are the same string, it returns true unless false. So the possible value is true or false. Our model is designed to capture different aspects of coreference. The model is also built to predict coreference for all of the candidate antecedents of a mention. This makes it useful for providing scores when the current mention has clear coreference links to many previous mentions. For example, ጠቅላይ ሚኒስትር ዐብይ might be linked to ጠቅላይ ሚኒስትሩ, ዐብይ, አብይ, ዐብይ

አህመድ, and ዶ/ር ዑበይ አህመድ::

Algorithm 4.5: Coreference Resolution Algorithm

Input : Amharic sentence

Output: Detected coreferring terms

Read Amharic sentence

For each Amharic sentence

Generate feature vector

Pass feature vector to classifier

Find antecedent- anaphor pair (i, j)

If classifier predicts i and j are coreferring

produce i and j

End If

End For

Trained Model

Model building is the process of creating/training new model with the help of model builder for this task based on the extracted features. In the proposed architecture the trained model is Coref model. This trained model is the final result of the training dataset and word embedding as well as feature extractor(learning algorithm).

4.4 Testing Phase

4.4.1 Preprocessing

Testing phase is another stage of the proposed model. Like training dataset, testing dataset requires preparation to be used as input during testing the model. This preparation is pre-processing step which has four stages. These stages are cleaning, tokenization and sentence segmentation. This preprocessing stage takes Amharic text as input. After we made pre-processing we test coref model with the help of coreference resolver.

4.4.2 Coreference Resolution

Coreference resolver is a system that clusters similar expression (mention pair) in the sentence. Mention pair detection is the process of identifying similar expressions and link them using clustering with corresponding compatible mentions in the text. So, mention can be determined by exact string match, head much and distance in the sentence. The mention are clustered by calculating the probability of mentions with the invocation of the trained model. Finally, the outputs of Amharic predicted coreference chains are displayed as follows:

{"predicted_clusters": [[[0, 3], [12, 13], [21, 22], [25, 26], [46, 47], [58, 60]], [[66, 66], [74, 74]]], "speakers": [["-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-"], ["-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-"], ["-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-"], ["-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-"], ["-", "-", "-", "-", "-", "-", "-", "-"], ["-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-", "-"]], "doc_key": "ge: (amarticle.xml); part 000", "sentences": [["ዶ/ር", "አብይ", "አህመድ", "አሊ", "የተወለዱት", "ነሐሴ", "9", "ቀን", "1968", "ዓ.ም", "ነው", "።"], ["ዓብይ", "አህመድ", "የኢትዮጵያ", "ፌደራላዊ", "ዴሞክራሲያዊ", "ሪፐብሊክ", "አራተኛው", "እና", "የአሁኑ", "ጠቅላይ", "ሚኒስትር", "ናቸው", "።"], ["ጠቅላይ", "ሚኒስትር", "በመስከረም", "30", "ቀን", "2012", "ዓ.ም", "ለኢትዮ-ኤርትራ", "እርቅ", "ላደረጉት", "ታላቅ", "አስተዋጽዖ", "የ 2019 ን", "የኖቤል", "የሰላም", "ሽልማት", "አሸንፈዋል", "።"], ["በጥቅምት", "2014", "ዓ.ም", "ዓብይ", "አህመድ", "ለሁለተኛ", "የ 5", "ዓመት", "የሥልጣን", "ዘመን", "በይፋ", "ቃለ", "መሐላ", "ፈጽመዋል", "።"], ["ጠቅላይ", "ሚኒስትር", "ዐቢይ", "ወደ", "ጦር", "ግንባር", "ዘመቱ", "።"], ["ኢትዮጵያ", "በንጉሠ", "ነገሥት", "ኃይለ", "ሥላሴ", "ጊዜ", "ታዋቂ", "ከሆኑ", "አገሮች", "አንዷ", "ነበረች", "።"]]}

This predicted cluster includes the clustering numbers, the mentions, the speakers, and the document key with document name. Accordingly, the first number [0,3] indicates that "ዶ/ር", "አብይ", "አህመድ" and "አሊ" from the first sentence. The second number [12, 13] shows that "ዓብይ" and "አህመድ" from the second sentence. Therefore, [0,3] and [12, 13] are cluster in one group to show that both terms are coreferent. ዓብይ አህመድ is coreferent with ዶ/ር አብይ አህመድ አሊ. Generally [0,3] and [12, 13] is exactly predicted ዶ/ር አብይ አህመድ አሊ and ዓብይ አህመድ respectively as coreferential. The remaining clustering number continue like this. [21, 22]= ጠቅላይ ሚኒስትር from the second sentence and [25, 26]= ጠቅላይ ሚኒስትር from the third sentence. So, [21, 22] and [25, 26] are coreferents.

CHAPTER 5: EXPERIMENTATION AND EVALUATION RESULT

5.1 Introduction

This chapter presents the experimental aspects of the study. We explore the development tools and programming language that are needed to our study. We also discuss the evaluation result of our research work. The over all this chapter includes the experiment, data gathering, data preparation, development tools used to implement the proposed model and evaluation used to test the performance of the Amharic coreference resolution model. An experiment is the tangible evidence of all researchs that is very important for all researchers in various filed of specialization. Thus, the performance of the experiment should be evaluated by various metrics to know the performance of an algorithm for a given system. In experimental research, evaluation is nothing but the way of computing the abilities or quality of a developed system. The evolution of an experiment has an important role in Amharic coreference resolution system and other researchers. For example, it helps us to know the performance of our system and to check it with previous research work. For other researchers also offer them with the essential information that they need and they can easily compare alternative developed systems to pick up the one that come across their requirements. As such we explore various hyper-parameters results that are needed for our experiment. The next sub-section is experimental settings for a system.

5.2 Experimental Settings

Experiment of the proposed system is conducted on hp Laptop that contains Intel core i3, 1TB of internal hard disk and 4 GB RAM properties that are running on Ubuntu 16.04 operating system. The experiment started by installing Tensorflow and other libraries. The installed libraries are tensorflow-gpu, tensorflow-hub, h5py, nltk, pyhocon, scipy, sklearn, tqdm and colorama. Trained model of Amharic coreference system using BERT is took more than three days. Training a model using this approach is somehow it requires more time, but it produces better result when we compare with other technique like Long Short Term Memory (LSTM). Even if it takes more time to train the model, it produces better accuracy result.

5.3 Amharic Coreference Resolution System Development and Model Training

In this, we will present how prototype of the Amharic coreference resolution system and its components are developed based on the proposed research design. In the development of Amharic coreference resolution system a number of components is prepared effectively. These components are Amharic corpus, trained model, text preprocessing module that is capable of accepting Amharic text as input and so on.

5.4 Data Collection and Preparation

In this experiment we collect Amharic text from various reliable sources. In the proposed system the experimentation begins by collecting Amharic text that are feed into pre-process step. These corpus is collected from the Amharic bible, books, fictions, news agencies, broadcasting media, Wikipedia, and magazines. The collected documents are from diverse domains such as BBC News Amharic, business fields, politics field, religion, technology, health, science, history, entertainment, and so on. So, these collected documents are not homogeneous. The grasp Amharic text are used for word embedding, training and testing the proposed system. Two kind of collected document are used for training purposes. From this the first corpus is unlabelled Amharic text which is used for word vector generation and the second corpus is labelled Amharic text which is used for training and testing purpose. There is no official and good quality prepared dataset that are publicly available for the Amharic coreference resolution system during training and testing the model. There is no freely available dataset for Amharic coreference resolution. Therefore, Amharic coreference resolution dataset is organised/prepared in our study at the first time. This dataset is named as Amharic coreference resolution dataset. It shares some concept from Computational Natural Language Learning(CoNLL-2012) Shared Task [19].

The corpus is a collection of grammatically correct Amharic sentences. It is prepared in CoNLL-2012 format and then converts each sentence in to '.jsonlines' format which is used to train the model. The '.jsonlines' file format contains tokens of a sentence, speaker information and coreference identification number. After we convert the conll-2012 file format into '.jsonlines' file format the sentence is split in to tokens. Finally, we set-up two types of window size experiment such as 511 and 129. Thus, the higher value of windows size performing better result, but it takes more time during in bertification process (feature extraction). Access to the desired corpus in some languages like Amharic is difficult.

Data Format

Since coreference resolution task requires linguistic expression in the sentence. In this study we have prepared Aramaic datasets which is demarcated by "#begin document" and "#end document". In the document, the information of each text is arranged vertically with one word/token per line, and a blank line after the last word/token of each text. The prepared Amharic dataset has 3 main parts, these parts are described below.

1. ID/Document ID- It is the first column that identifies token in the sentence.

2. Token/Words – The second column is word forms of a sentence.

3. Coref/Coreference-This column shows that coreference relations that is noted in open-close notation.

Consider this sentence in order to understand part of the prepared Amharic dataset in Amharic coreference resolution using BERT.

ዶ/ር ዓብይ አህመድ አሊ የተወለዱት ነሐሴ 9 ቀን 1968 ዓ.ም ነው። ዓብይ አህመድ የኢትዮጵያ ፌደራላዊ ዴሞክራሲያዊ ሪፐብሊክ አራተኛው እና የአሁኑ ጠቅላይ ሚኒስትር ናቸው።

So the .conll file format begins with ‘#begin document’ and ends with ‘#end document’ as well as key value is ‘ge’ and document name is amharicarticle.xml; part 000. Look at this figure.

```
#begin document ge:(amharicarticle.xml); part 000
0	ዶ/ር	(0
1	አብይ	-
2	አህመድ	-
3	አሊ	0)
4	የተወለዱት	-
5	ነሐሴ	-
6	9	-
7	ቀን	-
8	1968	-
9	ዓ.ም	-
10	ነው	-
11	።	-

0	ዓብይ	(0
1	አህመድ	0)
2	የኢትዮጵያ	-
3	ፌደራላዊ	-
4	ዴሞክራሲያዊ	-
5	ሪፐብሊክ	-
6	አራተኛው	-
7	እና	-
8	የአሁኑ	-
9	ጠቅላይ	(0
10	ሚኒስትር	0)
11	ናቸው	-
12	።	-

#end document
```

Table 5.1: Format of the Coreference Annotations

ID	Tokens	coref
1	ጠቅላይ	(1
2	ሚኒስትር	1)
3	ዐቢይ	(1)
4	ወደ	-
5	ጦር	-
6	ግንባር	-
7	ዘመቱ	-
8	።	-
1	ዶ/ር	(1
2	አብይ	1)
3	ኢትዮጵያን	(2)
4	ይወዳል	-
5	።	-

As the table shows that the last column is the coreference resolution number column that follows an open-close representation with a mention number in parentheses. Every word has an ID number in the sentence, and every mention is marked with the coref number of the entity it refers to. An opening bracket shows that the beginning of the target mention, while a closing bracket shows the end of the the mention.

5.5 Development Tools and Programming Languages

In the development of this research, we used various development tools. These tools are Tensorflow-gpu, Tensorflow-hub, h5py, nltk, pyhocon, scipy, sklearn, tqdm, colorama,…etc.

Tensorflow: It is designed by the Google team for various task. This framework is an open source machine learning library that includes python API's to implement machine learning and deep learning models in the easiest way. This library has a lot of abstraction power that users working with money computational primitive wrappers such as matrix operations, element-wise maths operators, and looping control. It links together the computational algebra of optimization methods for easy computation of different mathematical expressions. We preferred Tensorflow for this research because of its many feature such as optimizes and calculates mathematical expressions easily with the help of multi-dimensional arrays. In our experiment, we used this framework to train and build a model for Amharic coreference resolution and word embedding in easiest way. In this study, Tensorflow 1.13.1 is installed to explore our experiment that used as a back-end for Keras library.

Scikit learn: Scikit-learn (Sklearn) is another library which is used for coreference task. So this is the most advantageous and key library to machine learning approach in various research that is implement using python programming language. It provides a number of tasks for machine learning and statistical modelling such as clustering, classification and so on. This library also used in a wide range of state of the art deep learning algorithms for supervised and unsupervised problems. It is very easy to use with high performance result in machine learning activities. Beyond this Sklearn is used to evaluate both machine learning classifiers and deep neural classifiers by calculating performance metrics of a model.

NLTK: The Natural Language Toolkit (NLTK) is the platform used to analysis and process human language data. This toolkit applies in statistical NLP task which has text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning functionalities. We used this for tokenization.

Pyhocon

It is a parser for Python programming language that is used to convert any HOCON content into '.jsonlines' file format.

Notepad++: It is a free available tool which is used in our study to prepare the Amharic corpus for training and testing and used to edit python code.

Wendershare EdrawMax: We used this tool to draw the architecture of Amharic coreference resolution model.

Python 2.7 Programming Language: We installed Python 2.7 to implement the proposed Amharic coreference resolution model. It is the most powerful and general purpose programming language which has a verity of applications from scientific and mathematical calculations to desktop graphical user interfaces.

Hornmorpho: This is the developed tool for Amharic, oromo and Tigrinya language. It is built using python programming language. The main aim of this tool is to analyse the morphology of three Ethiopian languages such as Amharic, Oromo, and Tigrinya [6]. We have used this tool to get morpheme information of words.

Hyperparameter Settings: Hyper-parameters have some parameter values which are used to control the training process in machine and deep learning. It is a configuration settings to build a model in various learning approach. We implement that parameter with various value of dropout rate, learning rate, batch size…etc. The hyper-parameter setting that we used to train Amharic coreference resolution model is concluded in Table 5.2.

Table 5.2: Hyper-parameter Settings

Model hyper-parameters	Value
filter_widths	[3, 4, 5]
filter_size	50
char_embedding_size	8
contextualization_size	200
contextualization_layers	3
ffnn_size	150
ffnn_depth	5
feature_size	20
max_span_width	30
use_metadata	true
use_features	true
model_heads	true
coref_depth	2
lm_layers	4
lm_size	1024
coarse_to_fine	true
refinement_sharing	False

Learning hyper-parameters	Values
max_gradient_norm	5.0
lstm_dropout_rate	0.4
lexical_dropout_rate	0.5
dropout_rate	0.2
optimizer	adam
learning_rate	0.001
decay_rate	1.0
decay_frequency	100
ema_decay	0.9999

Learning Rate

The value of learning rate can affect the train process of the model. A low value of learning rate will slows down the learning process of the model. It counts slowly and it will take many times. In opposite to this, when the value of learning rate is larger it can speed up the training process in neural networks. Generally, it is hyper-parameter that is used in the training of a model and it has a positive value between 0.0 and 1.0. In our study we used 0.001 value of learning rate to build Amharic coreference resolution model using BERT.

Hidden Layers: In our study BERT uses 12 layers of transformers block with a hidden size of 768 and number of self-attention heads as 12. To construct our model the following configuration are used.

Table 5.3: Configuration Settings

NO	Parameter	Descripion
1	vocab_size:	Vocabulary size of inputs_ids in the model.
2	hidden_size:	Size of the encoder layers and the pooler layer.
3	num_hidden_layers	Number of hidden layers in the Transformer encoder.
4	num_attention_heads	Number of attention heads for each attention layer in the Transformer encoder.
5	intermediate_size	The size of the "intermediate" (i.e., feed-forward) layer in the Transformer encoder.
6	hidden_act	The non-linear activation function (function or string) in the encoder and pooler.
7	hidden_dropout_prob	The dropout probability for all fully connected layers in the embedding, encoder, and pooler.
8	attention_probs_dropout_prob	The dropout ratio for the attention probabilities.
9	max_position_embeddings	The maximum sequence length that this model might ever be used with. Typically set this to something large just in case (e.g., 512 or 1024 or 2048).
10	type_vocab_size	The vocabulary size of the token_type_ids passed into the model

5.6 Evaluation

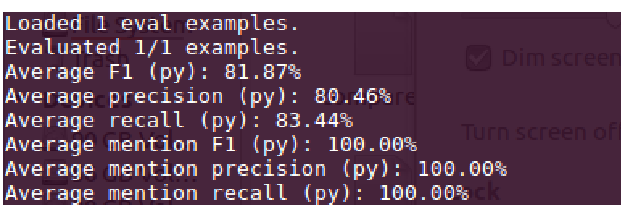
The major difficult task in Amharic coreference resolution system is about detecting the number mentions that are available in the sentence. Even if there is ambiguity about the number of existing mentions in the text, we evaluated the accuracy of the model by standard evaluation metrics of coreference resolution. We have run our experiment on the Amharic corpus.

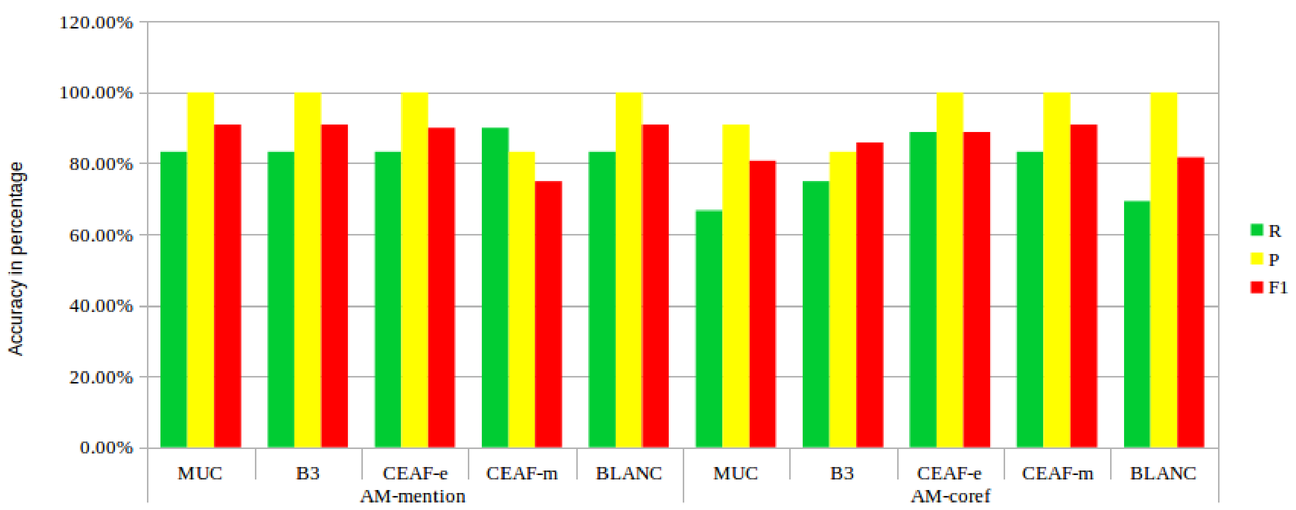
We have trained the proposed method and evaluate the experiment with the MUC, B^3, CEAF-e, CEAF-m, BLANC and CoNLL metrics. To do this, we have a set of gold entities and a set of response entities. Gold entities are commonly referred to as key entities in the coreference resolution. Response entities are those entities that are generated by a coreference resolver. The experiment result is presented in the table below.

Table 5.4: Experiment Result

	Metrics														
	MUC			B^3			CEAF-e			CEAF-m			BLANC		
Model	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1
AM-mention	83.3%	100%	90.9%	83.3%	100%	90.9%	83.3%	100%	90.9%	83.3%	100%	90.9%	83.3%	100%	90.9%
AM-coref	66.6%	100%	80%	75%	100%	85.7%	88.8%	88.8%	88.8%	83.3%	100%	90.9%	69.3%	100%	81.7%

Sample screenshot of the precision, recall and F1 result of the model.





Evaluation result in all metrics for a given model

5.7 Discussion

We explore an experiment for Amharic coreference resolution as shown the result of the experiment above. In the experiment first we collect and prepare Amharic text dataset for Amharic coreference resolution. After we collected the dataset, we manually prepared annotated a corpus. Then, we apply feature extraction from mentions and building word embedding for semantic and syntactic features of the word relation [38].

An experiment of Amharic coreference resolution system using BERT is conducted by achieving better performance result. The value of result are 81.87%, 80.46% and 83.44% of F1, precision and recall metric respectively. Yaregal Assabie and Temesigen Dawit evaluated the experiment of Amharic anaphora resolution system using 10-fold cross validation technique [11]. Based on the collected dataset, they achieved a success rate of 81.79% for resolution of hidden anaphors whereas an accuracy of 70.91% is obtained for resolution of independent anaphors [11].

CHAPTER SIX: CONCLUSION AND RECOMMENDATION

6.1 Conclusion

In this research work, Amharic coreference resolution model is developed for Amharic language using bidirectional encoder representation from transformer. Thus, this chapter presents conclusion, contribution of the work, recommendations and future work. The main aim of this study is to design and measuring coreference resolution for Amharic text using transformer-based model. Coreference resolution is the process of clustering mentions that refers to the same real-world entity in the text. Coreference resolution is very important in NLP applications such as machine translation, question answering, text summarization and so on. In this paper, we propose a transformer based model, named BERT, for Amharic coreference resolution task.

Amharic corefernce resolution model has various parts such as training and testing phase as well as Amharic predicted mention. The training phase has the following part: preprocessing (Cleaning, sentence segmentation and tokenization), word embedding, feature extraction and coref model. The testing phase also has preprocessing steps and coreference resolution. BERT helps us to study contextual representations of words that are assigned to sub-words in each word. BERT can be used in various NLP tasks such as question answering, name entity recognition, text classification, coreference resolution and so on. This model is highly expressive language model.

To test the performance of the model, we have used different evaluation metrics. The best performing system obtained F-scores of 80%, 85.71%, 90.9%, 88.86% and 81.7% on the test set with MUC, B^3, BLANC, CEAF-m and CEAF-e evaluation metrics, respectively.

6.2 Contribution of the Thesis

The main contributions of this thesis work are summarized as follows:

- We proposed new architecture for Amharic coreference resolution model.
- We prepared Amharic coreference resolution dataset, which is used as input for our model.
- We successfully use the BERT model for Amharic coreference resolution and obtained significant result.
- We developed word embedding for Amharic.
- We adapt algorithms from others to develop Amharic coreference resolution system, because a language has unique syntax and features.
- We experimented our model with the Amharic corpus which is created on this study.

6.3 Future Work

Coreference resolution task is very complex task which needs knowledge like morphology, syntactic, semantic, world knowledge and other type of knowledge. Applying of all this type of knowledge is time consuming and needs an intensive effort. As a result, we've concentrated on coreference resolution for Amharic text using bidirectional encoder representation from transformer (BERT). Since transformer based mode requires a huge amount of training data to get better result. So the performance of the model is maximized by adding the number of training dataset. Using a small n umber of data can affect the accuracy of Amharic coreference resolution model (coref model). Some of the recommendations that we suggest to be done in the future to increase the performance of the model is listed below.

- If you use a syntactic tree you can reduce noisy spans and it allows us to increase the number of antecedent candidates to capture long distance coreferent pairs up to one page.
- Amharic documents which are POS tagged and chunked are not found in different source. Therefore, we recommend that prepare POS tagging documents using tools to get high performance.

REFERENCES

[1] ባዬ ይማም, “የአማርኛ ሰዋስው,” አዲስ አበባ፤ የትምህርት መሳሪያዎች ማምረቻና ማከፋፈያ ድርጅት ማተሚያ ቤት, 1987.

[2] Webster, “A Balanced Corpus of Gendered Ambiguous Pronouns,” *rans. Assoc. Comput. Linguist*, vol. 6, pp. 605–617, 2018.

[3] D. Z. and Sandra K, “A language-independent system for coreference resolution,” *In Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 96–108, 2010.

[4] Veena. G. and Deepa Gupta, “A Learning Method for Coreference Resolution using Semantic Role Labeling Features,” *International Conference on Advances in Computing, Communications and Informatics*, pp. 978-1-5090-6367–3/17/1EEE, 2017.

[5] W. S. and H.Ng, “A Machine Learning Approach to Coreference Resolution of Noun Phrases,” *Computational Linguistics*, vol. (27) (4), pp. 521–544, 2001.

[6] Weter. M. and Hate G., “Coreference resolution,” *Computational Linguistics*, vol. 27(4), pp. 521–544, 2001.

[7] M. V. and J. Burger, “A model-theoretic coreference scoring scheme,” *in Proceeding of 6th Conference on Message Understanding, Columbia*, pp. 45–52, 1995.

[8] R. M. and Richard Evans, “A new, fully automatic version of Mitkov’s knowledge-poor pronoun resolution method,” *in Proceedings of the third International conference on Intelligent Text Processing and Computational Linguistics*, p. pp 17-23, 2002.

[9] D. L. and Tim Miller, “A pronoun Anaphora Resolution System based on Factorial Hidden Markov Models,” *n Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland*, pp. 1169–1178, 2011.

[10] A. B. and Breck Baldwin, “Algorithms for scoring coreference chains,” *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, pp. 563–566, 1998.

[11] Temesigen Dawit and Yaregal Assabie, “Amharic anaphora resolution using Knowledge-poor approach,” *Springer International Publishing Switzerland*, pp. 278–289, 2014.

[12] D. Jur. afsky and James H. Martin, “An introduction to natural language processing, computational linguistics, and speech recognition,” 2007.

[13] S. M. and Lamia Hadrich, “Anaphora in Arabic Language: Developing a Corpora Annotating Tool for Anaphoric Links,” *the international Arab conference on information technology*, 2008.

[14] Gabi Rolih, “Applying Coreference Resolution for Usage in Dialogue Systems,” *Uppsala University*, 2018.

[15] Suresb, A., “BERT for Coreference Resolution,” *In Proceedings of the Conference of the German Chapter of the Association for Computational Linguistics*, pp. 59–74, 2019.

[16] Devlin, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171–4186, 2019.

[17] J. D. and Ming-Wei, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *Computing Research Repository*.

[18] M. R. and E. Hovy, “BLANC: Implementing the Rand index for coreference evaluation,” *Natural Language Engineering*, vol. 17, pp. 483–491, 2011.

[19] S. P. and Alessandro Moschitti, “Conll- 2012 shared task: Modeling multilingual unrestricted coreference in ontonotes,” *In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Conference on Computational Natural Language Learning – Shared Task*, pp. 1–40, 2012.

[20] T. S. and Ionov M., “Coreference Resolution for Russian: The Impact of Semantic Features,” *Computational Linguistics and Intellectual Technologies. International Conference “Dialog 2017” Proceedings*, pp. 339–349, 2017.

[21] A. W. and Pierre Nugues, “Coreference resolution for Swedish and German using distant supervision,” *Gothenburg, sweeden*, pp. 23–25, 2017.

[22] A. H. and D. Klein, “Coreference resolution in a modular entity-centered model, in Human Language Technologies,” *The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California*, pp. 385–393, 2010.

[23] H.L. and Khoa N. A, “Coreference Resolution in Full Text Articles with BERT and Syntax-based Mention Filtering,” pp. 157–171, 2006.

[24] B. K. and Amir Globerson, “Coreference resolution with entity equalization,” *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 673–677, 2019.

[25] L. M. and Marta Recasens, “Coreference resolution: an empirical study based on semeval-2010 shared task 1,” *University of Spain*, 2010.

[26] A. Bj and orkelund, “Data driven multilingual coreference resolution using resolver stacking,” pp. 49–55, 2012.

[27] K. C. and Christopher D, “Deep reinforcement learning for mention-ranking coreference models,” *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2256–2262, 2016.

[28] Heeyoung Lee, “Deterministic coreference resolution based on entity-centric, precision-ranked rules,” *Computational Linguistics*, vol. 39(4), pp. 885–916.

[29] A. S. and Sabine Weber, “Different German and English coreference resolution models for multi-domain content curation Scenarios,” pp. 48–61, 2017.

[30] D. K. and Greg Durrett, “Easy victories and uphill battles in coreference resolution. In Proceedings of Empirical Methods in Natural Language Processing,” 2013.

[31] V. S. and Jason Eisner, “Easy-first Coreference Resolution,” *Mumbai*, pp. 2519–2534, 2012.

[32] Lee He and L. K., “End-to-end neural coreference resolution,” *In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017.

[33] K. C. and Christopher D., “Entity-centric coreference resolution with model stacking,” *In Association for Computational Linguistics (ACL)*, pp. 1345–1359, 2015.

[34] R. W. Bruno Oberle, “French coreference for spoken and written language,” pp. 1261–1273, 2019.

[35] Jieyu Zhao, “Gender bias in coreference resolution: evaluation and debasing methods,” 2018.

[36] Zhao, “Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods,” *In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, pp. 16–23, 2018.

[37] J. Z. Tianlu Wang, “Gender bias in coreference resolution: Evaluation and debiasing methods,” *n Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, pp. 15–20, 2018.

[38] Pennington, “Global vectors for word representation,” *In Proceedings of the conference on empirical methods in natural language processing (EMNLP), Doha*, pp. 1532–1543, 2014.

[39] Pradheep Elango, “Hybrid methods for coreference resolution in Swedish,” *Stockholm University*.

[40] C. K. Manning C., “Improving Coreference Resolution by Learning Entity-Level Distributed Representations,” *Association for Computational Linguistics Proceedings*, vol. 1, pp. 643–653, 2016.

[41] N. Cardie, C. V., “Improving machine learning approaches to coreference resolution,” *In Proceedings of ACL, Philadelphia, Pennsylvania, USA*, pp. 104–111, 2002.

[42] S. W. Alexander M, “Learning global features for coreference resolution,” *In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 994–104, 2016.

[43] Jheng-Long W. and Wei-Yun M., “A Deep Learning Framework for Coreference Resolution Based on Convolutional Neural Network,” *Taipei, Taiwan*, 2017.

[44] A. B. Jonas Kuhn, “Learning structured perceptrons for coreference resolution with latent antecedents and non-local features,” *In Association of Computational Linguistics (ACL)*, 2014.

[45] Vincent Ng, “Machine learning for entity coreference resolution: a retrospective look at two decades of research,” *University of Texas at Dallas Richardson, TX 75083-0688*, 2017.

[46] C. C. Wagstaff K., “Noun Phrase Coreference as Clustering,” *In Proceedings of the Joint Conference on Empirical Methods in NLP and Very Large Corpora*, 1999.

[47] R. K. K. van Deemter, “On co-referring: coreference in MUC and related annotation schemes,” *Computational Linguistics*, vol. 26(4), 63 629.

[48] Luo, X., “On coreference resolution performance metrics,” *In Proceedings of the conference on human language technology and empirical methods in natural language processing*, pp. 25–32, 2005.

[49] K. Andrew, “Probabilistic coreference in information extraction,” *In proceedings of the second conference on empirical methods in Natural language processing*, pp. 163–173, 1997.

[50] H. H. K. Min, “Pronominal Anaphora Resolution using a shallow meaning representation of sentences,” *Proceedings of the 8th Pacific Rim International Conference on AI*, pp. 862–871, 2004.

[51] Kovriguina, “Russian Tagging and Dependency Parsing Models for Stanford CoreNLP Natural Language Toolkit,” *International Conference on Knowledge Engineering and the Semantic Web*, pp. 101–111, 2017.

[52] H. D. I. John Langford, “Search-based structured prediction. Machine Learning,” vol. 75(3), pp. 297–325, 2009.

[53] A. H. Dan Klein, “Simple coreference resolution with rich syntactic and semantic features,” *Association for Computational Linguistics*, pp. 1152–1161, 2009.

[54] A. K. Vincent Ng, “Supervised models for coreference resolution,” *proceedings of the 2009 conference on empirical methods in natural language processing ACL and AFNLP*, pp. 968–977, 2009.

[55] A. K. Gerhard Wohlgenannt, “The Implementation of the Mention-Ranking Approach to Coreference Resolution in Russian,” vol. 2, pp. 20–34, 2018.

[56] Rocha, G., “Towards a Mention-Pair Model for Coreference Resolution in Portuguese,” *In EPIA Conference on Artificial Intelligence; Springer: Cham, Switzerland*, vol. 10423, pp. 855–86, 2017.

[57] Mazurowski, “Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance,” vol. 21, pp. 427–436, 2008.

[58] E. B. Dan Roth, “Understanding the value of features for coreference resolution,” *University of lllinois, Urbana, IL 61801*, 2008.

[59] ጌታሁን አማረ, “ዘመናዊ የአማርኛ ሰዋስው በቀላል አቀራረብ,” አዲስ አበባ፤ አልፋ አታሚዎች። *(10 ኛ እትም)*, 2002.

[60] ጌታሁን አማረ, “ዘመናዊ የአማርኛ ሰዋስው በቀላል አቀራረብ።,” *Addis Ababa*, *Ethiopia*, ንግዶ ማተሚያ ቤት, 1989.

[61] ባዬ ይማም, “የአማርኛ ሰዋስው,” *(የተሻሻለ 2 ኛ እትም)*. አዲስ አበባ፤ ብርሀንና ሰላም ማተሚያ ቤት, 2000.

Appendix A: Sample Amharic Coreference Resolution Dataset for Training

<COREF ID="0" TYPE="IDENT">ኃይሌ ገብረ ሥላሴ </COREF>ሚያዝያ 10 ቀን 1965 ዓ.ም <COREF ID="1"
TYPE="IDENT">በአሰላ ከተማ</COREF> እርሲ ክፍለ ሀገር ተወለደ። <COREF ID="0" TYPE="IDENT">ሯጭ
ኃይሌ</COREF> የሚኖረው <COREF ID="2" TYPE="IDENT"> ኢትዮጵያ </COREF> ውስጥ ነው። ኃይሌ <COREF
ID="2" TYPE="IDENT">አገሩን</COREF >በጣም ይወዳል። <COREF ID="0" TYPE="IDENT">ኃይሌ ገብረ
ሥላሴ</COREF> እጅግ ያልተገኘለት ድንቅ የረጅም ርቀት <COREF ID="0" TYPE="IDENT">ኢትዮጵያዊ ሯጭ</
COREF> ነው። <COREF ID="0" TYPE="IDENT"> ኃይሌ </COREF> በሩጫ ዘመኑ በ10 ሺህ, በ 5 ሺህ በግማሽ
ማራቶን, በማራቶንና እንዲሁም በሌሎቹ የሩጫ ዓይነቶች ከ11 በላይ የዓለም ክብረ-ወሰን ሰብሯል። <COREF ID="0"
TYPE="IDENT"> እሱም </COREF>በሩጫ ችሎታው ጠንካራ የሆነበት ምናልባትም <COREF ID="1"
TYPE="IDENT"> ከመንደሩ </COREF> ከባሕር ጠለል በላይ ከፍታ ነው ተብሎ ቢታመንም የተፈጥሮ ጥንካሬው ዓለምን
ያስደነቀ ብርቅዬ <COREF ID="0" TYPE="IDENT">ኢትዮጵያዊ</COREF> ነው።

Declaration

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

Declared by:

Name: Lingerew Bantie Asmare

Signature: ____

Date: ____

Confirmed by advisor:

Name: Yaregal Assabie (*PhD*)

Signature: ____

Date: ____