Chapter – 1
INTRODUCTION

If you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart.

– Nelson Mandela

According to a data created by electronic media, Guajarati language is spoken by 46.1 million people, all over the world. It is an Indo-European language and ranks 25th. This fact suggests the importance and weightage of the language.

It has its own script, descended from Devnagri.

Its Lingua sphere is spread mainly in western part of India that is known as Gujarat and Daman, Diu, Dadra and Nagar Haveli (India).

As a linguistic feature, Gujarati language has changed from time to time. Basically Gujarati belongs to Sanskrit. Gradually Sanskrit has changed into Prakrit, and Prakrit into Apbhransh. Apabhansh turns into old Gujarati and now we are using modern Gujarati. Grammar describes the language. It is the most important part of every language. Gujarati grammar has also played an important role in describing the language. One of the pioneering books in Gujarati grammar is ‘Sidhdha-Hem Shabadnushasan’ by Hemchandrachary.

SCOPE
This era is known as an electronic era. The means of communication are also based on electronic devices, for example, mobile phone, computer etc... Through these devices we can communicate all over the world. But sometimes language becomes a barrier. Many languages are spoken all over the world. Human beings cannot understand every language of world. Our thinking and thought process is expressed in our mother tongue very nicely. This is the biggest problem in India. All the people can not understand English very well. And for that reason they are not able to connect various devices.
PROBLEM
For this reason many linguists thought that a grammar should be prepared to make machine understand the language, so that the language barrier can be removed. Computational Linguistics is one of the subjects through which we can undertake the task of Machine Translation.

Many people from various disciplines try to translate with the help of machine. As a native speaker of Gujarati, I thought of providing small contribution in this task. The title of my thesis is ‘The Analysis of kAraka Relations of Gujarati, (with a view to use in NLP applications)’.

The sentence is divided into two different parts. A noun phrase and a verb phrase. kArakAs (theta roles) prepare the noun phrase, so, the area of my work is on the noun phrase. Every noun has relations with the verb, and these relations are known as kArakAs or theta roles.

KArakAs (theta roles) are semantic entities. They are represented through the case suffixes on syntactic level. They are known as Theta roles in modern linguistics. They are related to the verb and these relations are called as kArakAs in Indian Grammatical Tradition.

In this work I will give the analysis of the whole noun phrase. In the noun phrase, a Noun is the main component. Other components are Pronouns, Adjectives, Conjuncts and Suffixes like Gender, Number and Case, and Postpositions.

In this thesis I will try to give a sample of how machine automatically analyzes of Gujarati noun phrase.

METHODOLOGY
For machine it is not easy to identify nouns, pronouns, adjectives, conjuncts and suffixes like gender, number and case and postpositions. So, rules are made for the identification of nouns and other components and so through these rules machine can try to identify and analyze.

All the grammatical categories are discussed in detail in chapter number three. I have used tree diagrams to analyze these grammatical constituents.
Through these tree diagrams I have tried to show different kArakaAs (Theta roles) and noun phrases which occur in a sentence. Some noun phrases are also a part of the verb phrase. All nouns are related with verb. These nouns have their own roles in the situation denoted by the verb. These roles are known as kArakaAs (Theta roles) in Indian grammatical tradition.

As a native speaker of Gujarati, I have taken all the examples from regular conversations, and a few books of Gujarati.

With the help of these examples and through their tree diagrams I have given a grammatical analysis.

From this analysis few rules for machine are generated. Nouns are divided into six types on semantic basis. But since computational analysis can not work on semantic basis, they are classified into two different major categories, nouns which take covert gender and nouns which take overt gender. Then again for machine it was divided into three types, Common noun, Proper noun, and Verbal noun. These classes are useful but not computationally recognizable hence, sample of nouns are listed into three classes. Different paradigm tables are prepared accordingly. These paradigm tables help a machine to analyze nouns.

Gujarati sentences are analyzed through tree diagrams. Then through the help of paradigm tables machine will give an analysis of nouns.

Here in this thesis a few examples will be represented for the analysis of machine.

CHAPTERIZATION

First chapter deals with the general introduction of the thesis. In this chapter the central idea of work is given and an importance of grammar is shown. How it can be used in NLP is also shown.

The second chapter describes the sentence. When we talk about language a sentence plays an important role to produce a single thought. Further more a sentence is divided into two different parts. A noun phrase and a verb phrase. A noun phrase is made up of different components like Nouns, Pronouns,
Suffixes of Gender, Number and Case, Postpositions, Conjunctions, Interjections, and Adjectives, and in verb phrase Verb root, Auxiliary and Suffixes of Gender, Number, Person, Tense, Aspect, and Mood, and Adverb. All these are known as parts of speech.

In the third chapter, noun phrase is described in detail. First section deals with the types of nouns in Gujarati. Grammatically they are divided into six different types. But when we give classification in terms of computation they become three. They denote kArakAs (theta roles).

A Pronoun and its different components are described in detail. Nouns and Pronouns take Suffixes like Gender, Number and Case. Case suffixes play a major role to show relations between noun and verb. These relations are known as kArakAs (theta roles). kArakAs are semantic entity of grammar. Case suffixes are syntactic entity of grammar. Number shows the difference between one and more than one. Gender is arbitrary in Gujarati. But all nouns have Gender and they are divided into three class. They are Masculine, Feminine and Neuter.

Postpositions play a major role like case suffixes. Basically they are nouns and written with a space in Gujarati.

An adjective is defined as a word which gives more information about noun or pronoun. Adjectives describe nouns in terms of qualities like size, shape, color, number etc... All these things are discussed in detail in this chapter.

The fourth chapter is based on kArakAs (theta roles). This is the main chapter in my thesis. These kArakAs are part of noun phrase. There are thirteen different kArakAs (theta roles) in Gujarati. Basically all of them are nouns. That is why the analysis of the Noun phrase is undertaken. The kArakAs are the semantic category. They are realized through case suffixes on syntactic level. They denote the relation between a noun and the verb in the sentence.

In the fifth chapter the verb phrase is described. The kArakAs denote the relation between the noun and the verb. So the verb phrase is described in detail.
The types of verb roots, auxiliary verbs, suffixes like Tense, Aspect and Mood and Gender, Number, Person, are discussed in detail.

In the sixth chapter the Computational part is discussed. Machine Translation is a part of computational linguistics. Machine Translation is a long process. From that long process the analysis of nouns along with their kAraka relations is done. All the kArakAs are nouns, and these nouns are divided into three major categories. The paradigm tables are prepared for the analysis of these nouns.

Last chapter is the conclusion. This is a very vast field, but I have tried to give a sample analysis of kAraka relations in a sentence. A CD of Software is also given to see the applications of the analysis.

**APPENDIX**

At the end of the thesis the appendix is given in which the lists of nouns are given. Some of the possible outputs are also given with examples. The paradigms of nouns are prepared on the morpho-phonemic bases. Some of the paradigms are also given in the Appendix.