Chapter-1
Introduction

The aim of this research is to study the dynamics of the English word *of* and to find its functional equivalents in Telugu for the purpose of Machine Translation. The overall approach of the investigation is resolutely descriptive and empirical. It is, mostly, an approach driven by translation needs.

This research is mostly related to English-Telugu machine translation system. Here I have adopted some of the popular machine translation methods: Rule Based Method, Example/Parallel Corpora Based Methods and Dictionary Based Method. Thus, there will be a visible influence of these methods on whatever I discuss in this thesis. Mostly, we look at the input sentences to render them using sense of these methods to get most natural translation. For example, to avoid the unnecessary analysis, idiomatic sentences will be put in the example or database list, which is one of the methods of machine translation.

This thesis contains seven chapters. First chapter is about methodology and the history of the research on *of*. Second chapter is about the grammar of the English word *of* and its Telugu correspondences. Third chapter discusses the distribution of *of* in English. Fourth chapter is about the data and its analysis. Fifth chapter is about the translation process found after the data analysis. Sixth chapter is about the implementation. And seventh chapter is conclusion.

1.1. Introduction:

In the process of translation, we have to work on two issues: First one would be, an explicit analysis of the source language in terms of structural elements and their
meaning, etc. and the second one would be finding the functional equivalents to these in the target language. The quality of the translation depends on the resolution of the issues and obtaining acceptable equivalents.

Some categories of the source language may not be available in the target language. For instance some languages, like English, may have prepositions and some other may not. Languages like Japanese, Telugu, etc do not have the concept of preposition. In Japanese nouns take postpositions, according to the function, i.e. subject, object, etc. The Telugu language doesn’t have the category preposition, but have a whole set of corresponding postpositions, for example gūrci/guriñci, nuñci, etc. While translating such words into the target language the source language’s category may have its corresponding category rather than the same category. Though Telugu doesn’t have prepositions, but it has post positions which serve more or less similar function.

Consider prepositions, one of the closed class syntactic categories of English, they are polysemous and homonymous in nature. However, words, especially of, found in various word combinations offering various degrees of mapping problems in the target language. Most of the time they are grammaticised and don’t have explicit lexical status.

1.2. Methodology:

1.2.1. Identification of the problem:

i. No original meaning exists: Until a few hundred years ago of and off were merely different pronunciations of the same word. In Old English of is an unstressed form of ‘æf’ (preposition, adverb) with the meaning of ‘away’ and ‘away from’. Primary sense in Old English was still ‘away’, but it was later used to translate the Latin genitive case and so acquired all the meanings expressed in Latin by the genitive. And it has become the most grammaticised word.


6
ii. Most frequent word: In a corpus study, independent of domain, of is the second most frequent word in English. 100 million corpus of BNC has 6, 187, 267 entries of the word the as the most frequent word and with 2, 941, 444 entries of stood second. In a written corpus which contained 89.7 million words, the word the occurred in 5, 776, 384 times, and of occurred in 2, 789, 403 times. Here also of is the second most frequent word. In both corpora and is the third most frequent word.

iii. Special identity: Because of the word’s frequency and it’s almost exclusively postnominal function, CLAWS (Constituent Likelihood Automatic Word-tagging System – developed by the Center for Computer Corpus Research on Language, Lancaster University) has given a separate tag for of, PRF, whereas all other prepositions are given a common tag PRP.

iv. Found in many categories: Of comes with different combinations of syntactic categories in English like phrasal verbs, complex prepositions, phrasal nouns etc.

v. Most grammaticised: Complement in the structure of NPs, complement in the structure of AdjPs and AdvPs.

1.2.2. The Data:

1000 specific sentences containing of were taken from the Cambridge Advanced Learner’s Dictionary (the dictionary contains a total of 39,332 sentences with of), Cambridge University Press 2003. These sentences contain various constructions involving nearly 150 nouns, 81 phrasal verbs, 351 phrases and idioms, and some other combinations.

These constructions are cross checked with the BNC corpus and some examples sentences are taken from it.

1.2.3. Translation:

The above mentioned sentences involving phrasal nouns, phrasal verbs and idioms and other combinations were translated into Telugu for the purpose of comparison
and analysis. The translations are analyzed and will be listed out the explicit functional equivalents for the word of are listed.

1.2.4. Generalization:
On the basis of these translations some generalizations are made.

1.2.5. Implementation:
The generalizations in the form of rules are implemented in a machine translations system, that is being developed and tested at the Language Technology Laboratory, CALTS, University of Hyderabad

1.1.3. Earlier works on of:
So far, as far as our knowledge goes the word of has not been studied explicitly as an independent item and we have not been able to find any relevant literature from the point of translation. However, different syntactic categories which contain of have been studied, like complex prepositions, phrasal verbs, idioms etc. Some of the details these works discussed in the following section.

Some of the information used in the following section is from secondary sources. (And the most of the data that researchers used (according to them) for diachronic study are computerized historic corpora, Helsinki Corpus, CALD, BNC)

**Complex prepositions:** It is difficult to date the first reference to complex prepositions in the literature, the preposition-like nature of PNP-constructions had certainly been recognized by a number of writers by the second half of the nineteenth century.

Matzner (1873), for example, discusses ‘Prepositionale Formeln’ (prepositional formulae’) and restricts his description to those constructions in which the second prepositional element is of. Matzner’s list thus covers only a subset of the sequences
commonly known today as complex prepositions. He hasn’t given them full prepositional status but only states that they ‘resemble prepositions’.

According to Hoffmann (2005), in the course of time the syntactic usage has been changed as in the typical example *instead of*, which has its roots in the Old English *in stede* plus a genitival NP: \[ [in stede] of X ] \(\Rightarrow\) \[ [in stede of]X \]. The formerly concrete locative reference of *in stede* + NP (people and objects) has gradually extended to carry a predominantly substitutive meaning that in middle English could also be applied to abstract activities. Today, the Germanic word *stede* meaning *place* is almost exclusively used in *homestead* or in the archaic *in his/her/its stead*. Thus, *stead* has clearly lost its syntactic and lexical independence and *instead of* is a fixed and indivisible phrase. Finally, spelling is another indication of this gradual process with today’s standard two-word spelling first occurring in the 16th century. It may be said that most of the complex prepositions of modern English are formed like this.

Sweet (1892) takes a more sophisticated approach in that he clearly highlights the distinction between form and function. He describes complex prepositions as ‘consisting of a noun governed by a preceding preposition and followed by another preposition, which grammatically governs the following noun, although logically the noun is governed by the whole group’. Structure and function thus do not coincide: while PNP-constructions are described structurally as consisting of individual parts, the entire combination is described as functioning as a single preposition.

Givon (1993) makes this distinction even clearer. In his discussion of *pseudo-*possessives, he notes the ‘discrepancy between surface form and semantic-grammatical reality’ in complex locative and offers the following examples and explanation:

a) True possessive:

He measured the front of the house.

It was 30 feet wide. (‘it’=the front)

b) Pseudo-possessive:
He stood in front of the house.
It was 30 feet wide. (‘it’=the house)

The contrast in the above examples indicates that a semantic and grammatical historical reanalysis has taken place in expressions such as (b). The reanalysis pertained to which noun is the head of the NP and which one is the modifier. In (a), the original possessive modifier construction indeed retains its original semantic status. In (b), the historical reanalysis has conspired to enrich the inventory of locative prepositions in English, giving rise to new complex prepositions.

Givon’s analysis makes it clear that complex prepositions must be examined within a historical context. A number of earlier grammarians also take a historical perspective and try to trace the development that led to the establishment of complex prepositions. Among them is Earle (1892), who writes that “through the phrasal prepositions we are able to see how the older prepositions came into their place, and (to speak generally) how the symbolic element sustains itself and preserves itself from decay by inanition a presentive word gets enclosed between two prepositions, as if it had been swallowed by them, and were gradually undergoing the process of assimilation. By and by the substantive becomes obsolete elsewhere, and lives on her won as a preposition, with a purely symbolic power.” (Earle, 1892:515). As an illustrative example of this process, he offers the development of in despite of to despite. Thus for Earle complex prepositions are only an intermediate stage on the way towards (new) single-word prepositions.

According to Kruisinga (1932) in the whole range of PNP-constructions, many of which would certainly not be considered prototypical complex prepositions by today’s grammarians: in presence of, for purpose of, beyond reach of, beyond hope of and under cover of. Kruisinga goes on to note that:

“In all these cases, the construction with the article (in the presence of, etc.) is also possible and perhaps equally current. The reason is that each element of the group has retained its independent meaning in the minds of many speakers. But there are
outwardly identical cases when the group has so much unity that no article is ever used. This applies to ‘by reason of, by way of, in virtue of, in imitation of, in case of, in hopes of, by dint of, etc”. (Kruisinga 1932: 345).

The borderline between complex prepositions and free sequences of adjacent PP phrases is not easy to determine. No distinctive formal or functional features can be defined and a closer look at standard grammars suggests that lists of complex prepositions are based on intuition rather than precise grammatical criteria. Quirk, (1985:671-2) offer nine different criteria which help in the distinction between complex prepositions and other, freer combinations. For example, in spite of allows no modifications whatsoever (*in the spite of, *in spite to, *in spites of) whereas in pursuit of one can find in the pursuit of and in hot pursuit of etc. they establish a ‘scale of cohesiveness with in spite of at one extreme behaving in all possible ways like single word prepositions – and on the shelf by (the door) at the other extreme, allowing all nine types of variation. However, the authors stress that these ‘indicators of syntactic separateness’ form a gradient scale and consequently offer no opinion as to where the cut-off point should be drawn.

According to Hoffmann Sebastian (2005), not only the spelling and usage, the meaning also changes in many cases. Early newspaper texts show a completely different use from what can be found both in Gutenberg and the BNC (examples are given in other chapters).

Phrasal verbs: According to Tom McArthus, the term ‘phrasal verb’ was first used by Logan Pearsall Smith, in ‘Words and Idioms’ (1925), in which he states that Oxford English Dictionary Editor Henry Bradley suggested the term to him. Alternative terms for phrasal verb are ‘compound verb’, ‘verb-adverb combination’, ‘verb-particle construction’ etc. Many English verbs consist of two parts, a base verb and another small word. The small word is either a preposition or an adverbial particle. In some cases, the meaning of a two-part verb is simply a combination of the meanings of the two words, ex. come in, run away, walk across, sit on, etc. In some cases, the first word keeps its
meaning, but the second has a special intensifying sense. It means something like completely or thoroughly, Ex. *break up, tire out*. In other cases, the new two-part verb has quite a different meaning from the two separate parts, Ex. *give up* means surrender, and *blow up* means explode.

Other than the above, there are many other issues related to the *of* has to be studied explicitly for the purpose of machine translation.

In the next chapter, we will discuss the grammar points related to *of* in English. There, we will explore the basic syntactic units of English and their functions, besides, corresponding Telugu grammar will be discussed.