1.1. Aim and Scope of the Study

The nature of lexical entries of verbs has been the most sought after issue for research among Generative Morphologists. What is interesting in these works is the investigation with respect to representation and the role of argument structure in the linguistic description of the morphosyntax of the respective languages. The relationship between syntax and semantics can best be captured through investigations on argument structure and through various mechanisms underlying it. Besides attempting a partial description of the argument structure of Telugu verbs, this work probes the relevance of the syntactic valency to find what extent it can be predicted from the lexico-semantic representations associated with individual predicates in Telugu when they involve in derivation or compounding. One of the major goals of this study is to come up with a proposal and illustrate by a practical implementation of the argument structure to know that it is aptly relevant and it is very crucial in the disambiguation of different uses of the same verb form involving a number of senses. Though the registered title of the thesis is broad in its scope, it could not be maintained for various academic and non-academic reasons. Firstly, there is a sea change in the attitudes of people towards linguistics, and the goals of research in Applied Linguistics at the Centre during the last five years. While working at the Language Technology Laboratory of the Centre, I have been constantly reminded of the discussions centering around word sense disambiguation and particularly the verb sense disambiguation.
This has been the major problem in the development of machine aided translation systems (a number of such systems are being developed at the Centre). This has led to the change in the focus of the thesis. This dissertation centers around the verb sense disambiguation using argument structure. One of the chapters which focuses on this issue is Chapter-5 –“Argument Structure and Verb Sense Disambiguation in Telugu: A Computational Implementation”. The basic idea in writing this thesis is to develop a prototype application of verb sense disambiguation where argument structure figures as the main issue. The thesis is an outcome of various efforts in understanding the theoretical concepts underlying the argument structure, understanding the argument structure of Telugu verbs, mainly the representation of the argument structure and the computational implementation and testing. The study reports here at least two tangible results, viz. a near exhaustive study of the argument structure of Telugu verbs and a tool for computer applications involving verb sense disambiguation in Telugu. This thesis does not claim to be a contribution to the theory of argument structure directly or indirectly but it can claim to be a precaution in the development of Natural Language processing tools in the area of word sense disambiguation involving argument structure.

1.2. Significance of the study

The grammatical information of a lexical entry of a predicate may be analyzed through various ways viz. semantic structure, argument structure, grammatical function structure, and grammatical category structure. In this work, valency of predicates in terms of argument slots
and their semantically determined relative prominence has also been attempted. Semantic patterns of arguments are captured through thematic roles. This information may be expressed in a variety of ways; appealing directly to grammatical functions such as subject and object; (as in Lexical Functional Grammar (Bresnan 1996) or Relational Grammar (Blake 1990)), or to syntactic configuration (as in Principles and Parameters Theory (Chomsky 1981)), or to some combination of grammatical functions and category labels (as in Head-Driven Phrase Structure Grammar (HPSG, Pollard and Sag 1994)) (rf Louisa Sadler and Andrew Spencer. 1998) It is empirically tested that there is a semantic level of representation characterized by the argument structure of a verb and the specific properties of the arguments. Note that the marking of verbs exhaustively using linguistically standardized thematic roles is beyond the scope of this work. However, it is recognised that agent, patient, theme, experiencer, locative, instrumental, goal and source are the roles needed to be marked in each lexical entry. An exhaustive identification and marking of thematic roles requires a greater effort, and precision which is not the main goal of the thesis.

The present work, thus focuses on the identification of the semantic functions of the arguments i.e. the thematic roles assigned by a verb to its arguments and the way in which the relational semantics of the verb is represented at syntax level. Argument structure is the most crucial and relevant level of representation for verbs. Argument structure is manifested distributionally in syntactic alternations giving rise to differences in subcategorization frames or in the properties of the
arguments of a verb. The subcategorization frames within and across classes can disambiguate the usages of a verb with more than one sense.

1.3. Methodology

Through the work presented here, it is intended to clear the ground for a later larger scale attempt to develop a system for verb sense disambiguation based on argument structure. In other words, the intention of the work here to show by demonstrate that the argument structure of verbs can be profitably exploited to construct an application which should be part of a machine translation system and similar other natural language processing applications requiring word sense disambiguation. In a Telugu root word dictionary containing approximately 64,614 words, we have extracted 11,629 verbs and studied them for their meanings. From this list, verbs with more than one distinct sense have been extracted which numbered around 1,427 (12.27%). Again, these verbs were subjected to a critical scrutiny to eliminate such cases wherever the so-called multiple senses are because of the choice of the target language equivalent rather than the distinct meanings of the verb. For this purpose we have used a Telugu-Hindi anusaraka Machine Translation Dictionary developed at CALTS, University of Hyderabad (Electronic version, 1999) and Gwynn’s Telugu-English Dictionary (1991). Finally, screening a total of 1000 verbs which include simple monomorphemic underived plus derived through affixation and compounding remained. During the initial phase of the work, every such verb with multiple meanings were provided with argument Structure manually. Since every distinct meaning/sense of the
verb required a distinct but corresponding argument structure, in practice the number of verbs with distinct argument structure have increased to more than two fold. The thematic roles that we have selected for this purpose include mainly Agent, Patient, Theme, Experiencer, Goal, Source, Location and Instrument. It was found later that during the implementation of a verb sense disambiguation, not all of them play the same role. The first four Agent, Patient, Theme and Experiencer play a key role in verb's sense disambiguation. Verbs have also been categorized on the basis of their argument structure. I have also studied the criteria for the predictability of argument structure in verb alternation or derivation. Whenever a new verb is derived through the processes of affixation or compounding, it is not always possible to predict the verb's argument structure. There is not a single process in Telugu verb derivation which facilitates the prediction of the new verb's argument structure.

Eg. \( k\text{Alu} \) 'to be burnt', 'to be toasted'
\( k\text{Alcu} \) 'to burn', 'to fire(as a gun)', 'to toast'

\( \text{welu} \) 'to be floated' (as on water and air), 'to be decided'
\( \text{welcu} \) 'to lift (make less heaviour), to decide'

\( \text{wlru} \) 'to be relieved', 'to be resolved'
\( \text{wlrx} \) 'to relieve, to resolve', 'to arrange', 'to satisfy'
As illustrated above, many such derivations are not always regular—both in the alternation or in semantics. This forces us to list all such verbs in the dictionary rather than deriving them through morphology.

A dictionary containing 64,614 words of different categories with paradigmatic information is used to obtain the correct analysis of these nouns and verbs from the test sentences. I have selected about fifty verbs, each of which have more than one argument structure frame. In other words, ambiguous verbs are selected from the verb list along with their multiple argument structures. For each such frame one or two
exemplary sentences are constructed. From these sentences, nouns are extracted and listed in a dictionary with the necessary semantic features. In terms of +/- human, +/-animacy, +/-concrete, +/-combustible article, +/-edible, +/-body part etc. As part of the argument structure of the verbs, arguments are also provided with similar semantic features of ontological relevance.

Ex.

awamu, P(+h) ‘he, distant, masc'
AmeV, P(+h) ‘she, distant, fm
Aku, N(-a, +c) ‘leaf'
goda, N(-a, +c) ‘wall'
katteV, N(-a, +ca) ‘stick'
kalu, N(+bp) ‘leg'
noru, N(+bp) ‘mouth'
annaM, N(+ed) ‘cooked rice'
pamdu, N(+ed) ‘fruit'

After arriving at a fairly good number of verbs whose argument structure frames are more or less available, I took to the issue of developing a procedure for Verb Sense Disambiguation.

An algorithm which lays out a step by step procedure for the implementation and working of the system is constructed. As a part of this procedure, sentences containing ambiguous verbs are fed to the program. The program reads each word and calls a sophisticated Telugu
morphological analyzer, which analyses each word form and lists root/stem forms plus their morphological categories. Then the program picks up the verb and matches it with one of the argument structure frames, where it picks up the predetermined number of arguments, which will be later matched and conformed based on their semantic features from a dictionary of nouns for disambiguation by narrowing down on a specific argument structure frame.

1.4. Limitations and future work

Several importatnt topics related to argument structure, for reasons of focus and practical implementation, we have also side lined from including deverbal nominals and deverbal adjectives. The main aim of the present work is to investigate the argument structure of Telugu verbs and show by implementation that argument structure can be used efficiently to disambiguate verb senses. However, the present work does not address the problem of each and every verb in Telugu that have different senses exhaustively. It is assumed that a classification of lexical entries for verbs can be captured with the notion of an argument structure frame introduced in this work, if the thematic properties reflected in the alternations of argument structure and corresponding to the features of individual verb senses are predicted. The major limitation of this work is that the nouns in Telugu must be exhaustively analysed and marked for their semantic features. It is certainly a stupendous work but will have greater gains in the long run particularly in the area of Natural Language Processing.
1.5. An outline of the work

The work reported in this thesis is organized into six Chapters followed by a Reference and an Appendix; Chapter-1 is an Introduction where aim and scope of the thesis is stated and the significance of the work in the current scenario is discussed. The discussion on methodology regarding the development of database of verbs and nouns and their semantics referring to certain large and standard dictionaries, procedures for the implementation of the algorithm and testing are discussed. Towards the end, certain limitations of the current work and of the possible future work are mentioned Chapter two deals with Telugu Morphology and Syntax. It also examines the Telugu nominal structure, verb structure, finite verb agreement and word order which are relevant to the argument structure. The third chapter deals with the valency of verb and conceptual frame work that is necessary for grounding any study with reference to the argument structure. It is a study that touches the semantic structure, the argument structure and their inter relationships in terms of layers. This chapter provides the reader an over all view to situate the argument structure in terms of its functions and manifestations in syntax and semantics. Chapter four is a description of the preliminaries in the argument structure of Telugu verbs. It takes into consideration of the earlier works on Telugu. It starts with the standard definition of arguments and argument structure and other related areas like the classification of the verbs with reference to the argument structure and valency. The Fifth Chapter is the main chapter of the thesis and it is conceived as an exercise in computational implementation of argument structure to disambiguate verbs. It
discusses the need and the studies in word sense disambiguation in natural language processing. It then describes the model proposed based on an algorithm of implementation and testing. The chapter demonstrates that an application based on argument structure for the purpose of verb sense disambiguation can be built. The chapter VI is a concluding note on the thesis.