Prepositional Phrase Attachment and Generation of Semantic Relations

M.S. Final Stage Project Report

by

Ashish Almeida
Roll No: 03M05601

under the guidance of

Prof. Pushpak Bhattacharyya

Department of Computer Science and Engineering
Indian Institute of Technology Bombay
Mumbai, India
Abstract

This project largely deals with the prepositional phrase attachment problem and generation of semantic relations, which are difficult problems in the field of natural language processing. In this work we looked at it through the knowledge based perspective, i.e., by encoding the world knowledge and language phenomena in terms of rules and dictionary features. The work is more focused on solving the attachment problems of English sentences, thereby improving the accuracy of analysis process. The problem needs, particularly, a knowledge intensive solution. We have used insights from linguistics, towards solving this problem. We achieved good results based on our strategy of using ‘argument structure information and feature rich lexicon’ for prepositions ‘of’ and ‘to’. Also, the usefulness of automatic extraction of features for words in the dictionary becomes evident through the work.
Acknowledgments

I would like to thank my guide Prof. Pushpak Bhattacharyya for his regular guidance and constant encouragement. I would also like to thank Dr. Rajat Kumar Mohanty for his linguistic insights, with whom I closely work on this project. I would like to mention the informative discussions on linguistics, I had, with Ms. Debasri Chakrabarty. I thank all my colleagues at CFILT lab and NLP-AI group for their support at various stages of my project.

Ashish Almeida.
IIT Bombay.
December 2006.


## Contents

1 **Introduction** ........................................ 7
   1.1 Problem Definition ................................ 8
       1.1.1 Structural Ambiguity .................... 8
       1.1.2 Generation of Semantic Relations ....... 11
   1.2 Overview ........................................ 12

2 **Literature Survey** .................................. 16
   2.1 UNL and English Analysis .......................... 17
   2.2 Preposition Phrase Attachment Resolution .... 18
   2.3 Semantic Roles and Preposition Sense Disambiguation 21

3 **Introduction to Universal Networking Language** 23
   3.1 Universal Networking Language (UNL) ........... 23
       3.1.1 Syntax .................................... 24
       3.1.2 Universal Words (UWs) ................... 25
       3.1.3 Relations and Attributes ................ 26
   3.2 UNL System ..................................... 26
       3.2.1 Encoder .................................. 27
       3.2.2 Dictionary ................................. 29
       3.2.3 Rules ..................................... 30
       3.2.4 Analysis Process: an example ........... 31
       3.2.5 Deconverter and Generation Process ..... 32
4 Linguistic Foundations for Preposition Phrase Attachment
   4.1 Syntactic Frame ................................... 35
   4.2 Subcategorization Frame ............................ 36
      4.2.1 Significance of Subcategorization Frames ... 36
      4.2.2 Types of Subcategorization Frames ......... 38
   4.3 Selectional Restrictions .............................. 40
   4.4 Thematic Relations .................................. 41
      4.4.1 Application of Thematic Role ............ 42
   4.5 Arguments of Nouns .................................. 43
   4.6 Ontology ............................................. 43

5 Prepositional Phrase Attachment Strategy .............. 45
   5.1 Linguistic Analysis of Prepositional Phrases .... 46
      5.1.1 The Frame [V-NP1-P-NP2] .................... 47
   5.2 Implementation of Prepositional Phrase Attachment
      in UNL ................................................. 48
      5.2.1 Augmenting the Dictionary Entries with Syn-
            tactic and Semantic Features ............... 48
      5.2.2 Determination of UNL Relations ............. 49
      5.2.3 Design of the Rule Base to Handle Preposi-
            tional Phrases ................................. 50
   5.3 Evaluation Process .................................... 52

6 Processing of ‘of’ ........................................ 53
   6.1 Introduction .......................................... 53
   6.2 Distribution of of-NP Constructions in English ... 54
      6.2.1 The Frame [NP1-of-NP2] ....................... 54
      6.2.2 The Frame [V-NP1-of-NP2] .................... 56
      6.2.3 The Frame [V-NP1-of-NP2-of-NP3] ............ 57
      6.2.4 Conclusions from the Linguistic Analysis ... 58
6.3 Implementation of of-PP in UNL ............... 59
   6.3.1 Augmenting the Lexical Entries with Argument Structure Information ............... 59
   6.3.2 Implementation to Handle the of-PPs ........... 60
   6.3.3 The Strategy .................................... 61
6.4 Evaluation ............................................. 61

7 Processing of ‘to’ ........................................... 65
   7.1 Introduction ........................................... 65
   7.2 Detecting the Role of ‘to’ .......................... 67
   7.3 Analysis of to-prepositional Phrase ............... 67
   7.4 Analysis of to-infinitival Clause .................. 68
      7.4.1 The Attachment Issue ......................... 69
      7.4.2 PRO in to-infinitivals ....................... 69
   7.5 Dictionary ............................................. 72
      7.5.1 Acquisition of Subcategorization Frame Information ................................. 72
      7.5.2 Extracting PRO Information .................... 73
   7.6 Assigning Semantic Relations ...................... 75
      7.6.1 UNL Generation for to-PP ...................... 75
      7.6.2 UNL Generation for to-infinitival Clause .... 76
   7.7 Strategy ............................................... 77
   7.8 Evaluation ............................................. 77
      7.8.1 Preparation of Test Sentences ................ 78
      7.8.2 Creation of Dictionary Entries ............... 79
      7.8.3 Testing ........................................... 79
      7.8.4 Error Analysis ................................. 79

8 Analysis of Temporal Prepositional Phrases .................. 81
   8.1 Structure of Temporal Prepositions ............... 81
      8.1.1 Temporal Field ................................. 81
8.1.2 Temporal Prepositions .................................. 82
8.2 From Properties to Relations .............................. 84
  8.2.1 UNL Temporal Relations Inventory .................. 84
  8.2.2 The Mapping between Temporal Preposition
       Phrases and UNL Relations .............................. 85
  8.2.3 Lexical Entry of Prepositions ....................... 85
  8.2.4 Temporal Attributes ................................. 85
8.3 Formation of Analysis Rules ................................. 87
  8.3.1 General Strategy ...................................... 87
8.4 Evaluation .................................................. 89

9 Conclusion and Future Work .......................... 92
  9.1 Conclusion ............................................... 92
  9.2 Future Work ............................................. 93
Chapter 1

Introduction

Natural Language Processing encompasses well known tasks such as spell checking, translation from one language to another, search engines, summarizer, question answering, speech to text among many other. These can be broadly classified into two processes: getting the meaning out of text or analyzing the language, i.e., natural language understanding and to get text from the meaning, i.e., natural language generation. This is illustrated in the figure 1.1 below.

![Diagram of Natural Language Processing](image)

Figure 1.1: Two main activities in Natural Language Processing

This thesis deals with the first task, i.e., analyzing a language. It involves converting the sentences from natural language to its meaning representation. This work delves into many classical, yet, non-trivial problems of NLP including but not limited to sentence...
boundary detection, word sense disambiguation (WSD), name entity detection, prepositional phrase attachment and anaphora resolution. This work revolves around solving attachment ambiguities for various cases in detail, in context of obtaining meaning representation. We have used Universal Networking Language (UNL): an interlingua as medium of representing meaning.

In this chapter, we elaborate on the problem of attachment within English language and semantic representation of sentences. We end the chapter with the overview of the rest of the report.

1.1 Problem Definition

This problem involves attachment ambiguity and generation of semantic relations, given a English sentence. The next two sections explain each aspect in detail.

1.1.1 Structural Ambiguity

A word, phrase, sentence or other communication is called ambiguous if it can be interpreted in more than one way. If the ambiguity is because of a multiple meanings of a word, it is called lexical ambiguity.

One type of ambiguity, called structural ambiguity, arises due to more than one possible structure for the sentence. Consider the example 1.1 which has a structural ambiguity.

(1.1) Time flies like an arrow.

This sentence has two possible interpretations as given in figure 1.2a and 1.2b. Figure 1.2a means *flies* named ‘time’ loves an arrow, whereas the other interpretation in figure 1.2b is ‘time’ passes *like an arrow* (as fast as an arrow). In other words, there are two different
parse trees possible for the sentence 1.1. In this case, both the meanings of a sentence are semantically valid and acceptable. Such sentences are said to be inherently ambiguous. Even human being needs a context to select the appropriate meaning. But there is other kind of ambiguity which concerns NLP. Certain sentences are interpreted only in one way by human being but multiple parses of such sentences are possible for a machine. Moreover, a machine can not select the meaningful interpretation out of the given possible parses because of the lack of world knowledge.

![Two possible parse trees illustrating ambiguity in sentence (1.1)](image)

Figure 1.2: Two possible parse trees illustrating ambiguity in sentence (1.1)

**Attachment ambiguity**

This is a specific type of structural ambiguity in which a clause or a phrase has more than one possible association in the tree structure of the sentence of which it is a part. If the ambiguity is about the attachment of a clause then it is called clause attachment and if it is about attachment of prepositional phrase it is called prepositional phrase attachment. Depending on the site of attachment there are at least two possibilities, noun attachment or verb attachment.
Verb Attachment: Consider example 1.2 with prepositional phrase attachment ambiguity.

(1.2) He kept the book on the table.

The sentence can be parsed in two ways as shown in figure 1.3. The prepositional phrase ‘on the table’ is attached to the noun ‘the book’ making a noun phrase ‘the book on the table’ as depicted in figure 1.3a which is semantically incorrect. Because, if we replace the NP ‘the book on the table’ in the sentence with just ‘the book’ the sentence becomes incomplete. The other parsing (figure 1.3b) suggests that the prepositional phrase ‘on the table’ is attached to the main verb ‘kept’ of the sentence. That is, ‘the book’ is not directly associated with ‘the table’. However, ‘the book’ is related to ‘the table’ through the verb ‘kept’. Thus ‘on the table’ is attached to the verb ‘kept’.
Noun Attachment: Consider an example 1.3 which is syntactically similar to 1.2.

(1.3) He read the book on physics.

The prepositional phrase ‘on physics’ is attached to the noun ‘book’ as shown in figure 1.4a. The reading in figure 1.4b is incorrect. For a computer, to determine the attachment site (noun or verb) is difficult as it needs to know the semantics of it. As it is evident from the sentences 1.2 and 1.3, syntax is necessary but not sufficient to solve the ambiguity in the attachment.

1.1.2 Generation of Semantic Relations

This process involves converting English Sentence to its meaning representation. It is done using a artificial language called UNL or Universal Networking Language. It is an interlingua understandable
to computers. It represents the sentence in the form of concepts and their relations.

(1.4) He read the book on physics.

For example, the simplified semantic graph or UNL representation for 1.4 is shown in figure 1.5. It shows four concepts (ellipse) and three relations (arrows). Note here that in ‘modifier-of’ relation between ‘book’ and ‘physics’, correct semantic relation entails correct attachment. The UNL is formally introduced in chapter three.

1.2 Overview

This chapter has introduced the problem of attachment ambiguity and semantics generation in case of English sentences. Chapter two describes the related work on analysis system and preposition phrase attachment. Universal Networking Language (UNL), which is used throughout the report as medium of meaning representation, is discussed in depth in chapter three. It also familiarizes the reader with the process used to transform the English sentence to UNL. This work largely banks on the ideas obtained from linguistics. Chapter
four introduces reader with some of the essential concepts and ideas in linguistics used in this thesis.

Chapter five to chapter eight deals with different aspects of attachment problem and semantics generation. Chapter five presents the detailed analysis and strategy of solving the prepositional phrase attachment. Chapter six and seven gives detailed account of ‘of’ and ‘to’ prepositional phrases. It involves linguistic analysis and actual implementation techniques. Chapter seven also incorporates infinitival role of ‘to’. Chapter eight describes exhaustive study of temporal function of prepositional phrases. Chapter nine concludes the work and mentions a few future directions of work. Appendix A-C gives overview of the tools and interfaces developed to be used for creating various resources for the project. There are very few or no resources in case of Indian Languages. We attempt to bridge this gap with the help of these tools.

The dependency between chapters is shown in the dependency tree in figure 1.6.

Summary of the contributions made

1. We analyzed the behavior of temporal prepositional phrases and mapped temporal prepositional phrases into the UNL (semantic) relations.

2. We found the significance of semantic head detection in of-NPs and based on it, analyzed them to get the accurate UNL representation [11]. We were able to achieve 92% accuracy.

3. For prepositional phrase attachment, we proposed the strategy based on the argument structure information. We get 88% and 82% results in case of ‘of’ and ‘to’ prepositional phrases.
Figure 1.6: Dependency among chapters
4. We have done the detailed analysis of all possible roles of the preposition for certain prepositions [22].

5. We proposed appropriate UNL representation of the sentences with clauses and PRO: the empty pronominal [21].

6. We used the conventional lexical and other resources to get the features required for the machine readable dictionary. We tried to capture syntax information from English Wordnet and Oxford advanced learner’s dictionary.

7. We developed various NLP tools which were not available for Marathi and Hindi language. These include concordance search utilities, Wordnet interfaces, dictionary editing and viewing etc.
Chapter 2

Literature Survey

As discussed in the first chapter, for analyzing the language, we have used UNL as an interlingua. There has been some work on the English to UNL \(^1\) system which is discussed here in brief. Though a lot of work has been done at various levels of language analysis, not many have attempted to solve the language analysis problem in its entirety. This chapter will describe three main streams of work which are related to what we want to do. Those are as follows:

(2.1) 1. UNL and English Analysis: It describes linguistic and computational aspects of UNL based analysis.

2. Preposition Phrase Attachment Resolution: It describes PP attachment in the context of V-N-P-N frame. The same frame is adopted for most of the work in the thesis.

3. Semantic Roles and Preposition Sense Disambiguation: These works discuss issues concerning semantic role-labelling.

\(^1\text{UNL or Universal Networking Language is an interlingua and it is explained in next chapter}\)
2.1 UNL and English Analysis

The paper titled “Predicate preserving parsing” by Jignashu Parikh, Jagadish Khot, Shachi Dave and Pushpak Bhattacharyya [8] talks about the overall process of analyzing an English sentence into UNL. They have used a language independent parser called ‘Enconvertor’ to achieve this. This parser uses a dictionary and rule base to create UNL graph from the source sentence.

(2.2) Dr. Sen received Nobel prize in economics.

[Sen] received Nobel prize in economics
received Nobel prize [in economics]
received [Nobel prize]
[received]

Consider the sentence 2.2. Enconvertor processes this sentence by reducing the sentence (partial parsing) and thereby generating the graph nodes. During the process it preserves the predicate till the end of the processing as shown in example 2.2. This process is elaborated in the chapter on introduction to UNL. The paper explains how the dictionary attributes are used to disambiguate the word sense, how the part of speech is identified and how the two different parse structures are distinguished.

Another paper on “Interlingua based English-Hindi machine translation and language divergence” by Shachi Dave, Jignashu Parikh and Pushpak Bhattacharyya [14] discusses the divergence issues pertaining to English to Hindi language in context of UNL. It studies the language divergence between English and Hindi and its implication to machine translation in context of the Universal Networking Language (UNL).

The paper describes the knowledge rich approach towards translation, wherein the information is provided to the system through
dictionary and language specific rules. It discusses various linguistic phenomena in which the two languages diverge in lexical and structural choices. It gives examples for each phenomenon and explains how the UNL captures the phenomenon with the help of dictionary attributes and rules. It believes that by handling more phenomena, it is possible to increase the accuracy of the system. The dictionary is made richer by augmenting its entries with more semantic features. These features are then invoked through the rules and matched to process text. The paper “The Use of Lexical Semantics in Interlingual Machine Translation” by B. Dorr talks about another interlingual approach [23] towards machine translation. It is heavily based on Lexical Conceptual Structures. This idea is linguistically motivated and discussed by R. Jeckendoff in the book titled "Semantic Structures". The paper also talks about various divergence cases pertinent to English, Spanish and German bilingual translation for the Machine translation system called UNITRAN.

### 2.2 Preposition Phrase Attachment Resolution

“Statistical models for supervised prepositional phrase attachment” by Advait Ratnaparakhi discusses a classifier [20] for attachment solving. The experiment constructs a classifier which maps an instance of an ambiguous prepositional phrase ([V N₁ P N₂] (e.g. see the boy on the hill)) to either of the two classes *viz.* noun attachment and verb attachment. The unambiguous examples of noun and verb attachment are extracted from parser’s output with the help of heuristics from syntactic structures. The author assumes that the information in just the unambiguous attachment events can resolve the ambiguous attachment events of the test data. He claims the accuracy of 82% for English text in this paper.
The paper “Structural ambiguity and lexical relations” by Donald Hindle and Mats Rooth [24] discusses prepositional phrase attachment problem. The early approaches used right association (typically verb attachment) and minimal attachment (typically noun attachment) as strategy to solve prepositional attachment problem. The author feels that the lexical information could be used for deciding the attachment. They use Fidditch parser to get the partial syntactic representation (ambiguities are kept as it is). It uses Associated Press news stories (13 million words) corpus. The parser is used to extract the heads of the noun and verb phrases and the association between the words in the [V-N-P] syntactic frame. Then, they use the log likelihood ratio to calculate lexical association (LA) as follows.

A positive LA score indicates verb attachment and a negative LA score indicates noun attachment. An LA score of zero means they are equally likely. A interpolation technique is used when the frequency of pair of words was very less. Testing is done on nearly 1000 sentences selected randomly from the Associate Press news stories. The attachment was decided manually by the author for the reference in testing the output of the machine. The precision of 80% and recall of 80% is obtained from the lexical associations procedure whereas the human judges give 86% precision and 86% recall. The 86% recall is because of the various other constructs which occur in the sentence.

In “An unsupervised approach to prepositional phrase attachment using contextually similar words” by Patrick Pantel and Dekang Lin [18] discuss the use of prepositional phrase attachment without the help of word sense disambiguation. The classification model is based on unified framework for disambiguation tasks suggested by Roth. The paper formalizes the problem as detecting attachment
score for $N_2$ among the two choices: $V$ or $N_1$ attachment in $[V\, N_1\, P\, N_2]$ syntactic frame. Probabilistic score is assigned for Verb attachment ($V$-score) and Noun attachment ($N$-score). Contextually similar words are used to deal with the data scarcity. If the $V$-score is higher than $N$-score $N_2$ is attached to verb else to $N_1$. But this decision is overridden in case of ‘of’ where the noun attachment is always preferred. The test data was extracted from Penn Treebank corpus as used by Ratnaparkhi [20]. They report overall 84% accuracy.

To find contextually similar words in the case of $[V\, N_1\, P\, N_2]$, the procedure is as follows. A cohort of word $w$ is defined as set of words that are in same dependency relation as that of $w$. Similar words are fetched from corpus based thesaurus. Collocations are used to find words in (semantic) relation to a word. Then, for two-word pairs $w_1$, $w_2$ cohorts are found for $w_1$ and intersected with similar words of $w_1$, which are obtained from collocation database to get the contextually similar words for $w_1$. The important idea is that this method avoids the sense disambiguation task.

Another very important related work, “A rule-based approach to prepositional phrase attachment disambiguation” [5] by Eric Brill and Philip Resnik, has used transformation based error driven learning to learn the transformation rules (The idea is commonly used in the part of speech tagging technique). Penn Treebank is used as the training corpus. 4-tuples, of the form $[V\, N_1\, P\, N_2]$ are extracted and part of those is used as a true attachment cases. The learner learns in the form of the transformation rules given a sample data-set. It uses the noun classes obtained from WordNet to create more general rules and avoid data scarcity and over fitting. It gives accuracy of 81.8% in case of noun classes for learning.
2.3 Semantic Roles and Preposition Sense Disambiguation

In the paper “Application of WordNet to prepositional attachment” by [12], the author has used WordNet as the main source of knowledge for preposition sense disambiguation. The papers talks about the preposition ‘of’. It assumes that ‘of’-prepositional phrase always has local attachment. In addition to the semantic relations and the synsets, it also uses the information provided by WordNet such as textual glosses. The algorithm begins with creating classes of V-of-N and N-of-N triplets. It uses wordnet semantic relations to find similarity between two words. The verb argument frame information is used along with gloss to find the relation between two words. Also, these classes then help to disambiguate the sense of the word used in that prepositional phrase. For example, in the noun phrase “acquisition of a company” the word *company* is semantic object of the action denoted by the noun *acquisition*. This information is available in the gloss of its synset. Such information can also be used to disambiguate the sense of the word. *I.e.*, in case of word *acquisition*, in a gloss of only one synset of it, the word *company* is the object.

In the paper “Classification of preposition semantic roles using class-based lexical associations” [15], Tom and Janyce have attempted to classify or disambiguate the prepositional roles at semantic level. The author extracts annotated prepositional roles from Berkeley Frame-net and Penn TreeBank corpus. It learns the semantic roles using Naive Bays classifier on collocations. It uses the features such as POS tags of previous and next *n* words. Then in the limited window, it looks for collocations grouped by the sense of the given word. Also, in another set of experiments, they try to augment
the words occurring in collocation of given word with its hypernyms. The author shows that the semantic classes obtained from Wordnet will help in generalizing. But the author makes no distinction in words before the preposition and words after the prepositions during classification, though it is linguistically relevant.

On Frame-net dataset it achieves 68.5% accuracy and on Penn Treebank it achieves 78.5% accuracy in classification of semantic roles of the prepositions. The author cites ‘frame based roles and fine-grained-ness’ as a reason of the low accuracy over framenet dataset. They also show the use of cross validation using framenet and treebank datasets and thus shows the generalizing ability of class based collocations.

The work described here gives brief outline of the work done in the field of English analysis and prepositional phrase attachment in particular. The prepositional phrase attachment problem discussed in introduction is dealt with in next few chapters. Before that, chapter three discusses UNL and how English to UNL system works. Chapter four talks about some essential theory in linguistics.
Chapter 3

Introduction to Universal Networking Language

Analysis in the simplest terms can be described as understanding the natural language as human being does. Specifically it means, inputting the text and producing some kind of annotated text.

There can be many levels to language analysis, viz. lexical, syntactic, semantic, discourse and pragmatic. The analyser we are concerned with, is restricted to the first three levels, i.e., lexical, syntactic and semantic level, which will generate an intermediate language representation leading to machine translation. This chapter will introduce the concept of Universal Networking Language as an interlingua and explains its components. It will also give the details of how the analysis is carried out for English sentences.

3.1 Universal Networking Language (UNL)

UNL is an Interlingua for multilingual information access on the web. It is an electronic language for computers to express and exchange information. In UNL, the knowledge content of sentences
is expressed in a form of semantic net like structures called UNL expressions. In this section, we will explain how the UNL sentence is made and describe in detail its various components.

### 3.1.1 Syntax

Typically, in UNL, a sentence of natural language is represented as a set of UNL expressions. A UNL expression is composed of universal words (UW), relations and attributes. Consider sentence 3.1. Its UNL representation is given along with it. UNL of a sentence consists of a set of UNL relations written each on separate line. The relations are directional and binary. The relation is directed from first UW to the second UW. Figure 3.1 shows the UNL of the sentence graphically.

(3.1) John wrote a letter to Mary.

\[
\begin{align*}
\text{agt} & \langle \text{write}(\text{icl}>\text{do}).@\text{entry}.@\text{past}, \text{John}(\text{iof}>\text{person}) \rangle \\
\text{obj} & \langle \text{write}(\text{icl}>\text{do}).@\text{entry}.@\text{past}, \text{letter}(\text{icl}>\text{document}).@\text{indef} \rangle \\
\text{gol} & \langle \text{write}(\text{icl}>\text{do}).@\text{entry}.@\text{past}, \text{Mary}(\text{iof}>\text{person}) \rangle
\end{align*}
\]

![Figure 3.1: UNL representation of a sentence 3.1](image-url)
In figure 3.1, there are three arcs corresponding to three relations. The nodes \textit{John}(iof>person), \textit{Mary}(iof>person), \textit{letter}(icl>document) and \textit{write}(icl>do) are the Universal Words (UW) corresponding to the words \textit{John}, \textit{Mary}, \textit{letter} and \textit{wrote} in (3.1). These are the words with restrictions in parentheses for the purpose of denoting unique sense. The term \textit{icl} stands for ‘inclusion’ and \textit{iof} stands for ‘instance of’. UWs can be annotated with attributes like number, time, aspect \textit{etc.}, which provide further information about how the concept is being used in the specific sentence. The arrows labeled with \textit{agt} (agent), \textit{obj} (object) and \textit{ins} (instrument) are the relations.

UNL system also depends on a resource called UNL Knowledge Base which essentially works as world-knowledge repository. In each UNL expression, one universal words is related to the other universal word with a relation. Thus, the relation is always binary. The set of such UNL expressions can be realized as a directed acyclic graph. Also, it has compositionality feature which means a UNL graph can have other UNL graph as its node.

### 3.1.2 Universal Words (UWs)

Universal words (UWs) are words of UNL. They are made up of two parts, the word indicating particular concept and its restriction. The restriction part imparts unique meaning to the UNL. For example, a concept of ‘cell’ as a functional unit of all organisms is realized in UNL as \textit{cell}(icl>living thing). Here, the restriction \textit{living thing} makes the word ‘cell’ unique by differentiating it with the meaning of ‘the jail’ or ‘electrical battery’. There is also a concept of compositionality through which whole sentence is represented as a single UW in other UNL expression. Such UWs are called compound UWs. It is one of the strongest feature in UNL which enables us to realize the complex sentence in intuitive manner.
3.1.3 Relations and Attributes

The relation connects two UWs in UNL. Thus, acting like a functional word of the natural language. The relations are always binary and directional. There are total 46 relations and 87 attributes as mentioned in the UNL specifications [16]. Attributes are associated with UWs and encodes sentence level information. They always begin with @ character. For example, there are attributes to show the time of the sentence, number (plural or singular), definiteness of a noun, mood and focus in the sentence. There is a special attribute called @entry, its function is to tag the root node in the set of UNL expression. Consider the sentences below.

(3.2) John ate.

Here, ‘John’ is related to ‘ate’ with a relation called agent. That is, ‘John’ is agent of the action ‘eat’. The attribute @entry indicates that it is root node and @past indicates that the action took place in past. Thus, its UNL will be

\[
\text{agt(eat(icl>consume).@entry.@past, John(iof>person))}
\]

3.2 UNL System

The UNL framework consists of two subsystems: analysis and generation. The analysis module converts the natural language input sentence to UNL expressions and the generation module converts the UNL expressions to the natural language sentence. Figure 3.2 shows the process block diagram. The analysis and generation modules each uses a language independent tool which is driven by language specific rule-base and uses a language specific UW-dictionary. The output of analysis module is UNL-expressions which is input for the generation system. The analysis tool is called Enconverter and
the generation tool is called Deconverter. The scope of this project is restricted to the analysis part. In the following sections analysis process is explained in detail.

![Block diagram of the UNL System](image)

Figure 3.2: Block diagram of the UNL System

3.2.1 Enconverter

It is a language independent analyzer [17]. It does morphological, syntactic and semantic analysis synchronously by accessing a knowledge-rich dictionary and interpreting the analysis rules. At every step of the analysis, the rule base drives the Enconverter to perform tasks like completing the morphological analysis (e.g., analyzing *boy* and ’s in the word *boy’s*), combining two grammatical entities (e.g., ‘is’ and ‘working’) and generating a UNL relation (e.g., *agt* relation between ‘he’ and ‘is working’). Many rules are formed using context free grammar (CFG)-like segments, the productions of which help in clause delimitation, prepositional phrase attachment, part of speech (POS) disambiguation and so on.

It is essentially a Turing machine which has many heads: processing heads and context heads. The processing heads are also called
analysis windows and are two in number: the left analysis window (LAW) and the right analysis window (RAW). The context heads are also called condition windows and they are many in number. They surround the two analysis windows. The machine scans the input sentence word-by-word picking the words through dictionary. A window, in simplest terms, is a pointer (handle) to the word in a sentence. All windows move together one step ahead or back at a time. Once the word is picked up from dictionary, a node corresponding to the word is created. An analysis window can select a dictionary entry by matching the attributes specified in the rules with those of the dictionary entry whose head word is a word in the sentence at hand. Condition windows can only select the words by matching the attributes in the dictionary entry with the attributes specified in the rule. Depending on the windows position, priority and the attributes they possess, Enconverter picks up the appropriate rule and executes it. The rules have capability to move the windows to and fro over the input doing action such as creating UWs, relating two closely neighboring UWs (thereby eliminating one of them), adding attribute to the already present nodes under the windows. During the analysis, whenever a UNL relation is produced between two nodes, one of these nodes is deleted from the tape and is added as a child of the other node to the tree. Forming
the analysis rules for the Enconverter is similar to programming a sophisticated symbol processing machine.

### 3.2.2 Dictionary

Dictionary contains many entries, one entry per line. Each entry consists of the head words (of the specific language), its corresponding universal word, a set of attributes and a three character code. To analyze a sentence, all the words in a sentence must be present in dictionary with its appropriate senses. Consider the sentence 3.3.

(3.3) Misha went to the garden.

Given below are the entries for the words in the above sentence.

|Misha| {} | ‘Misha(iof>person)’’ (N,ANIMT,PRSNS,PHSCL) &lt;E,0,0&gt; |  
|went| {} | ‘go(icl&gt;do)’’ (VRB,VGA,PAST,VLTN,TMPL) &lt;E,0,0&gt; |  
|garden| {} | ‘garden(icl&gt;place)’’ (N,INANI,PHSCL,PLC) &lt;E,0,0&gt; |  
|the| {} | ‘the’’ (ART,THE) &lt;E,0,0&gt; |  
|to| {} | ‘to’’ (PREP,#_TO) &lt;E,0,0&gt; |  
|[.|} | ‘’period’’ (PUNCT) &lt;E,0,0&gt; |  

There are 5 dictionary entries corresponding to 5 tokens in the sentence. The syntax for an entry is

|Head word| {} | ‘UW’’  
|list of attributes separated by commas | &lt;X,num,num&gt;|

The attributes give information about its lexical nature, structure, syntactic environment and semantics. The triplet at the end gives language code, frequency and priority. For example, the second entry ‘went’ is the past form of ‘go’. The attributes say, it is verb (VRB), action verb (VOA), past form (PAST), action (ACT), volition (VLTN), temporal activity (TMPL).

(3.4) [went] ‘‘go(icl&gt;do)’’ (VRB,VGA,PAST,VLTN,TMPL) &lt;E,0,0&gt;
3.2.3 Rules

The rule base consists of rules separated by new lines. They are used by Enconverter to analyze a sentence. Rules are made up of condition windows, analysis windows and priority at the end. There are around five thousand rules in the rule base. Implementation of the rule base involves activities such as deciding rule priority, part of speech disambiguation, structural disambiguation, picking right sense of a word. Besides, there are rules to create relations, modify attributes and modify nodes. The syntax of a typical rule is given below.

\[ r(LCW_1)(LCW_2){LAW}\{RAW\}(RCW_1)(RCW_2)Pxx; \]

- \( r \) is the symbol to indicate the type of rule.
- \( LCW \) is left condition window and \( RCW \) is right condition windows.
  There can be many condition windows. \( LAW \) and \( RAW \) are left and right analysis windows.
  \( Pxx \) is the number preceded by ‘P’ which indicates priority and it ranges from 1 to 255. A greater number indicates a higher priority.

The syntax of condition windows is as follows.

\[ (attribute_1,attribute_2,...,attribute_n) \]

\( attribute_1 \) to \( attribute_n \) is list of attributes separated by coma which is used to match against the attributes of the word obtained from the dictionary. The syntax of the analysis window is shown below.

\[ \{attrb_1,attrb_2,...,attrb_n:+/-attrb_1,+/-attrb_2,...,+/-attrb_n:rel:\} \]

It has four subparts separated by ':=' symbol. First part is attribute list similar to condition window. Second part is a list of attributes with a +/- sign indicating if the attribute needs to be added or removed from the node under the window. Third field is a three character code which indicates the relation to be assigned from the node under it to the other node under second analysis window. The last field is left unused. Now, consider the sample rule below.

\[ >\{ART,THE:::\}\{N,ABS:+&@def::\}(PRE,OF)(BLK)P30; \]
This is a right node modification rule (indicated by ‘>’). It deletes the left node (under left analysis window) and modifies the right node (under right analysis window). As it can be inferred from the attributes, the left analysis window is over the word which is the article ‘the’ and the right analysis window is on the word which is a noun (N) and it represents some abstract thing (ABS). Also, it matches the next word as preposition ‘of’ followed by blank. The rule adds a new attribute @def to the noun following article ‘the’ and deletes the article ‘the’. P30 indicates priority 30. This rule is applied only when all the conditions are matched in all four windows and there is no other high priority rule matching the same pattern.

3.2.4 Analysis Process: an example

Now, as we have seen how the analysis is done, consider a sentence 3.5 below. This sentence has a clause in it. Its analysis is given below step by step.

(3.5) The boy who stays there went to school.

The processing steps are as follows:

1. The clause ‘who stays there’ starts with a relative pronoun and its end is decided by the system using the grammar. The system does not include ‘went’ in the subordinate clause, since there is a pattern like [WH-Word V₁ ADV V₂] in the rule in the grammar which says that the verb (V₂) is an indicator that the clause ended before it.

2. The system detects ‘there’ as an adverb of place from the lexical attributes and generates plc (place relation) with the main verb ‘stay’ of the subordinate clause. At this point ‘there’ is deleted. After that, ‘stay’ is related with ‘boy’ through the
aoj relation and gets deleted. At this point the analysis of the clause finishes.

3. ‘boy’ is now linked with the main verb ‘went’ of the main clause. In the same way, the agt relation is generated after deleting ‘boy’.

4. The main verb is then related with the prepositional phrase ‘to school’ to generate plt (indicating destination), taking into consideration the preposition ‘to’ and the noun ‘school’ (which has PLACE as a semantic attribute in the lexicon). ‘to’ and ‘school’ again are deleted. From ‘went’, ‘go(icl>move)’ is generated with the @entry attribute which indicates the main predicate of the sentence and the analysis process ends. The final set of UNL expressions for the sentence 3.5 is given in 3.6.

\[(3.6)\] agt(go(icl>move).@entry.@past, boy(icl>person))
plt(go(icl>move).@entry.@past, school(icl>institution))
aoj(stay(icl>be), boy(icl>person)) plc(stay(icl>be), there))

Thus, in above sentence the relative clause is encapsulated as aoj and plc relations. The main clause is represented with agt and plt relations. Note that the knowledge content of the preposition phrase ‘to school’ is encoded through plt (place-to) relation between the words ‘go’ and ‘school’ and the preposition ‘to’ is not used in the UNL.

### 3.2.5 Deconvertor and Generation Process

UNL to natural language generation process involves conversion of UNL expressions to sentences in target language. The generation process is also called deconversion. The language independent tool used for this purpose is called Deconvertor. The analysis from source language to UNL and then, generation of target language sentences from UNL makes up a machine translation process.
The Deconvertor uses a set of rules and a UW dictionary which maps to target language words, to accomplish this task (refer figure 3.2). It involves two fundamental tasks: syntax planning and morphology generation. Syntax planning consists of deciding the word order in target language sentence. The word order depends on the target language and the UNL relation in which they occur. Syntax planning also involves selection and insertion of appropriate function words into the sentence. Morphology generation deals with generating correct forms of the word depending on its gender, number and aspect. The structure of the Deconvertor is similar to that of Enconvertor. It has multiple heads which can move to and fro and inserts words into string of words to generate a sentence. Depending on the state of the windows, current set of UWs and UNL relation, it selects the rule and changes its state accordingly.
Chapter 4

Linguistic Foundations for Preposition Phrase Attachment

This chapter introduces the linguistic concepts pertinent to the current work. It focuses mainly on the theory of argument structure of a verb, selectional restrictions, thematic roles and a brief description on ontologies with respect to the English WordNet.

Before trying to understand the argument structure, we need to know a few basic ideas and become familiarize with the linguistic terms. The noun phrase which appears before a verb in the English sentences is called subject of the sentences. The noun phrase which appears after a verb is called object of the verb. Subject and object are syntactic notions. The object is also called direct object to differentiate it with oblique object i.e. the prepositional phrase/clause. The direct object is the nearest element to a verb if present. Sometimes a sentence might have more than two verbs as in case of sentences with to-infinitival clauses. In such a case the predicate (verb) of outer or main sentence is called matrix verb.
For example, consider the sentence 4.1.

(4.1) The old man promised Ram to buy a cycle.

Here, ‘the old man’ is the subject and ‘Ram’ is the object of the sentence. The sentence has two verbs ‘buy’ and ‘promise’. ‘Buy’ is the verb in the infinitival clause and ‘promise’ is the matrix verb.

### 4.1 Syntactic Frame

The term syntactic frame broadly refers to a parts of speech sequence in a sentence. We know that in English, noun as a subject normally precedes a verb as a predicate. One such syntactic frame for verb is a verb occurring after a noun (i.e. [N V]). For now, we will consider the general influence of the context, i.e., the frames, on the interpretation of the parts of speech. For example, paints can be used either as a plural noun or a simple present tense verb. The main issue here is to distinguish between paints as a noun and paints as a verb. It is important to know that the syntactic frames provide information on the part of speech sequence of words of a sentence or part of a sentence. It does not determine abstract semantic descriptions. Consider the examples (4.2) and (4.3).

(4.2) I like the quality of those paints.

(4.3) He paints the garage every two years.

In example 4.2 ‘paints’ occurs as a noun, while in example 4.3 ‘paints’ occurs as a verb. This difference is evident from the different syntactic frames in which they occur. In 4.2, the form ‘paints’ occurs after an adjective ‘those’. It is followed by the noun ‘paints’. The frame for 4.2 is [PRON V ART N P ADJ N]. In 4.3, the form ‘paints’ occurs after a personal pronoun, which itself occurs at the
beginning of a sentence. That position for a pronoun signals very strongly, although not absolutely, that it is the subject of the sentence. Sentence subjects are normally followed by verbs. ‘paints’ in 4.3, follows a subject; ‘paints’ is, therefore, a verb. The frame of 4.3 is [NP V NP NP].

4.2 Subcategorization Frame

In syntactic theory, The subcategorization frame of a word is described as the number and types of syntactic arguments that it co-occurs with, i.e., the number and kinds of other words that it selects when appearing in a sentence.

(4.4) They ate the cake.

Thus, in the sentence 4.4, ‘eat’ selects, or subcategorizes for ‘they’ and ‘cake’. These entities are called complements of the word which subcategorizes for it. Using this terminology, one can refer to information about the range of complements which a given word takes as subcategorization information. Subcategorization frames are generally referred to as contextual information since it specifies the linguistic context in which a given item can be used. The Subcategorization features are restrictions on the range of categories which a given item permits or requires as its complements. For example, in case of sentence 4.4, the verb restricts the subject to be only animate thing and the object to be only edible thing.

4.2.1 Significance of Subcategorization Frames

Consider the following examples.

(4.5) John won’t invite Mary.

(4.6) *John won’t come Mary.
In examples, 4.5 and 4.6 the verbs are \textit{invite} and \textit{come} respectively. In the sentence 4.5 the verb \textit{invite} results in a grammatical sentence when followed by an noun phrase (NP) whereas in the sentence 4.6, it seems that the verb \textit{come} does not allow an noun phrase to follow it. Hence, it can be said that a verb like \textit{invite} takes an noun phrase complement, whereas a verb like \textit{come} does not. There does not seem to be any general way from which one can predict whether a given verb does or does not take a following noun phrase. This is not dependent on the meaning as well. For example, \textit{wait} and \textit{await} are synonymous in one sense but they do not share the sentence frame. Consider sentences 4.7 and 4.8.

(4.7) I shall await your instruction.

(4.8) *I shall wait your instruction.

Thus, there does not seem to be any way in which one can predict the complement information. In human beings, the subcategorization information is known to be part of knowledge he has. Thus, it is inferred that this is an idiosyncratic property of that lexical unit. Thus, this information should be included in the lexicon.

The subcategorization frame of verb gives the information about whether a prepositional phrase (PP) is part of the subcategorization frame or not. Thus, it is easy to determine whether a prepositional phrase occurring after the verb in a sentence is a part of the subcategorization frame, \textit{i.e.}, complement or not. If the prepositional phrase is not subcategorized by the verb, it is potentially an adjunct or has a local attachment. Hence, subcategorization information provides a substantial help in solving the prepositional phrase attachment problem. This idea is used in the subsequent chapters.

Along the same lines, the nature of clause attachment can be understood. The subcategorization theory is equally applicable to
nouns and adjectives.

### 4.2.2 Types of Subcategorization Frames

Each lexical unit varies in terms of the complement range it subcategorizes. Different types of Subcategorization frames for the verbs are discussed below. There are certain verbs such as ‘come’ that appear without any complement. Thus, the Subcategorization frame (given in square brackets) for this type of verbs is given in 4.9.

\[(4.9)\] come: [ ]

This notation says that ‘come’ appears with a zero complement. There exist some verbs which can both occur with and without an NP complement. For example, consider sentences 4.10 and 4.11.

\[(4.10)\] John won’t help Mary.

\[(4.11)\] John won’t help.

Such verbs will be doubly subcategorized as both transitive and intransitive.

\[(4.12)\] help: [NP], [ ] or [(NP)]

Certain verbs allow two NPs as their complement. These verbs are known as ditransitive verbs. Consider sentences 4.13 and 4.14.

\[(4.13)\] John gave [Mary] [a book].

\[(4.14)\] The postman handed [him] [a parcel].

The Subcategorization frame for ‘give’ verbs will be specified as in (4.15).

\[(4.15)\] give: [NP][NP]
Prepositional Complements

Some verbs permit one or more PP complements in their subcategorization frame. There are different verbs that take different PP complements headed by different prepositions.

(4.16) I defer [to/*at\(^1\)/on/*by/*with your suggestion].

(4.17) John waited [for/*to/*after the taxi].

(4.18) He put the book [on/under/*after/*before the table].

Sentences 4.16, 4.17 and 4.18 brings forth the fact that the choice of the preposition as a head of a PP is also an idiosyncratic information along with the selection of a PP complement. This motivates us to add the preposition information to the lexical entries in the lexicon. With this motivation a verb like ‘put’ will be kept in the lexicon as in (4.19).

(4.19) put: [NP] [PP\(_{on}\)]

Other than verbs there are also nouns and adjectives that allow a PP complement, as shown in sentences 4.20 and 4.21.

(4.20) Mary is fond [of/*with/*to] John.

(4.21) development [of /on/*at/*with the project].

Thus, it is also essential to include this information for the nouns and adjectives.

Clausal Complements

Some verbs allow clauses in their complement range. Consider the example sentences 4.22, 4.23 and 4.24.

(4.22) I knew [that he would come].

\(^{1}\)the * symbol indicates incorrectness of the sentence
(4.23) He asked me [whether I was leaving].

(4.24) I imagine [that you must be tired].

Thus, the information of clausal complements should also be included in the lexicon.

### 4.3 Selectional Restrictions

When an item subcategorizes a complement belonging to a particular category, it is not usually the case that any expression belonging to the relevant category can function as a complement of the item concerned; on the contrary, there are generally clear restrictions on the choice of complements. For example, the verb ‘*murder*’ subcategorizes for an NP complement; and yet there are severe restrictions on the class of NPs which can function as its object.

(4.25) (a) The boy murdered [the man].

(b) *The boy murdered [the tree].

(c) *The boy murdered [the stone].

(d) *The boy murdered [the lion].

The example sentences 4.25a–d shows that merely, the relevant category restrictions are not sufficient restrictions, since in each example, ‘*murder*’ subcategorizes an NP complement but still examples 4.25b-d are not acceptable. They violate selectional restrictions. Selectional restrictions are semantic restrictions on the choice of expressions within a given category which can occupy a given sentence position. The two notions of subcategorization and selectional restrictions are clearly distinct: subcategorization refers to syntactic frame whereas selectional restrictions are purely semantic in nature. The verb ‘*murder*’ subcategorizes for an NP complement but from
the examples in 4.25, it is evident that ‘murder’ selects a human being as a complement.

Following this discussion, it seems that selectional restrictions are idiosyncratic properties of a lexical item and should be specified in the lexical entries, in addition to subcategorization information. Thus, the entry for the verb ‘murder’ will be as shown in 4.26.

(4.26)

murder: Subcategorization frame: [NP]

Selectional Restriction: \{ HUMAN HUMAN \}

4.4 Thematic Relations

Linguists have argued that each argument of a predicate bears a particular thematic role and that the set of thematic functions which arguments can fulfill are drawn from a highly restricted, finite, universal set. Thematic roles are also known as theta-roles, or \( \theta \)-roles. Some of the commonly assumed theta-roles are given in table 4.1. For each role an informal gloss, together with an illustrative example, is specified.

(4.27) John gave Mary a book.

Thus, in example 4.27, ‘John’ bears the theta-role AGENT to the verbal predicate ‘gave’, ‘Mary’ bears the role GOAL, and ‘a book’ bears the role THEME. Thematic roles enable us to capture the similarity between different but related usages of the same lexical item.

(4.28) John rolled the ball down the hill.

(4.29) The ball rolled down the hill.
The italicized expression has a different constituent structure status in example 4.28 and 4.29. In example 4.28 ‘the ball’ is the object of the verb but in 4.29, it is the subject. Intuitively, it plays the same role in both the sentences, as ‘the ball’ is the entity undergoing the motion. This role-identity can be captured by saying that in both the cases ‘the ball’ bears the same thematic role. To be more precise, ‘the ball’ has the theta-role THEME in both 4.28 and 4.29, since in both cases, it is the entity undergoing motion. This information is also included in the lexicon.

4.4.1 Application of Thematic Role

As discussed in the previous section, the thematic role assigned by a predicate (usually a verb) is a semantic property of that lexical item and it is also an idiosyncratic property of the lexical item, similar to subcategorization information. Thus thematic roles in terms of UNL relations are encoded into lexicon. Whenever the UNL expression
is generated, the dictionary is consulted to get the UNL relations associated with the particular entry.

4.5 Arguments of Nouns

Some type of nouns do behave similar to verbs, i.e., they take arguments the way verbs take [7]. For example, in 4.30, the noun ‘destruction’ takes ‘city’ as its object-argument. In 4.31, the noun ‘addition’ takes two argument, ‘a little salt’ as an object and ‘water’ as goal.

(4.30) The destruction of the city.

(4.31) The addition of a little salt to water makes it good conductor of electricity.

4.6 Ontology

An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is defined as a systematic account of Existence. In both computer science and information science, an ontology is a data model that represents a set of concepts within a domain and the relationships between those concepts. It is used to reason about the objects within that domain. In context of our work, ontologies are the hierarchies of concepts. They help in decision making, i.e., in case of deciding the meaning of the word or an attachment site of a phrase in the sentence. It acts as a knowledge base for the analysis work.

Ontologies encodes the world knowledge into itself. For example, for ‘dog’, it gives information such as, it is a canine, an animal, a kind of mammal, a living thing. For ‘knife’, it gives information as it is a cutting tool, an artifact, a physical object etc. English Wordnet
[6] has hierarchies such as hypernymy, meronymy. For a given word, these hierarchies can be consulted to find out its properties.

Also the hierarchical organization of such features facilitates the process of dictionary building. For example, ‘eat’ demands an object which need to have attribute called edible object. Thus this attribute is essential for all words which are edible. This information can be extracted easily from Wordnet. As all the edible things, at some level in the hierarchy, are associated with the node edible thing.
Chapter 5

Prepositional Phrase Attachment Strategy

In this chapter we will discuss the prepositional phrase attachment problem in specific frame and possible solution to it. At the end of the chapter we propose a strategy to attachment of prepositional phrases which will be applied to preposition ‘to’ and ‘of’ in subsequent chapters.

We have already seen the problem definition for prepositional phrase attachment in Chapter 1. In this chapter we will analyze the attachment ambiguity in the limited syntactic frame as discussed in Chapter 2. (Here, it is assumed that NP is simple noun phrase without any embedded clause or preposition phrase into it.) Given a sentence containing the frame [V-NP₁-P-NP₂], there are two fundamental problems.

(5.1) 1. To determine attachment site for the prepositional phrase (PP, \textit{i.e.,} P-NP₂) (\textit{i.e.,} whether PP attaches to V or NP₁)

2. To determine the semantic relation that the PP has with the word to which it attaches to.
Significance of choosing such limited syntactic frame is that it does away any influence of other linguistic phenomenon on the testing of this attachment problem. Also, we get a chance to compare our result with the others.

5.1 Linguistic Analysis of Prepositional Phrases

Prepositions are often termed as syntactic connecting words. However, they have syntactic as well as semantic specifications that are unique to them. The selection of a preposition is decided by the meaning of the syntactic elements that determine it. The meaning of the preposition, i.e., its semantic role depends on the word to which it attaches and the object it takes.

A preposition can occur in different syntactic environments. For example, the preposition for participates in eight different syntactic environments. In each environment, its meaning is determined by the preceding words and the following noun phrase. Example 5.2 illustrates this with examples. Of this, we will only consider sentences of type 5.2(d), i.e., sentences involving [V-NP1-P-NP2] frame to restrict the scope of the problem.

(5.2) a. The search for the policy is going on.
   b. The main channel for breaking the deadlock is the Airport Committee.
   c. He applied for a certificate.
   d. He [is reading [this book] [for his exam]].
   e. The Court jailed him for possessing a loaded gun.
   f. She is famous for her painting.
   g. They are responsible for providing services in such fields.
h. They have been prosecuted for allowing underage children into the theatre.

5.1.1 The Frame [V-NP₁-P-NP₂]

The complexity of the frame [V-NP₁-P-NP₂] involves two research issues as mentioned in the introduction. Now let us analyze the semantics of this syntactic frame and see how it determines the attachment site. Consider these examples which represent each category.

(5.3) a. He [forwarded [the mail] to John].
   b. She wore [[a green skirt] with the blouse].
   c. We received [[an invitation] to the wedding].
   d. I can’t easily give [[an answer] to the question].

In example 5.3(a), ‘to John’ is second argument of the verb ‘forward’. Thus ‘to John’ attaches to the verb. In example 5.3(b), the prepositional phrase, ‘with the blouse’ can attach to the noun phrase ‘a green skirt’ or to the verb ‘wore’. As the verb ‘wore’ takes only noun phrase as an argument and the noun ‘skirt’ too does not specify with-PP as an argument, the PP is attached to the nearest element, i.e., the noun ‘skirt’. In case 5.3(c), the verb ‘receive’ expects noun phrase as first arguments and does not expect to-PP as its second argument. On the other hand, the noun ‘invitation’ demands an to-prepositional phrase as its argument. To fulfill the demand of the noun ‘to the wedding’ is attached to the noun ‘invitation’. In case 5.3(d), the verb ‘give’ demands the to-PP as its second argument. Also, the noun ‘answer’ specifies to as its argument. In this example, both noun and verb specify the to-PP as its argument, the preference is given to the nearest element, i.e., the noun ‘answer’.
We have analyzed the different possibilities of attachment depending on demand of argument structure, in a frame \([V\text{-}NP_1\text{-}P\text{-}NP_2]\). In section 5.2, we shall translate this linguistic insight into an algorithm which, in turn, is implemented into the Enconverter system of English.

## 5.2 Implementation of Prepositional Phrase Attachment in UNL

In this section, we shall present the algorithm to find attachment in the case of the frame \([V\text{-}NP_1\text{-}P\text{-}NP_2]\), before that we need to introduce the argument structure information in terms of dictionary attributes. These attributes will then be fetched to make decisions based on it.

### 5.2.1 Augmenting the Dictionary Entries with Syntactic and Semantic Features

We have already observed that the argument structure of a lexical item plays an important role in determining the attachment site for an NP in a frame like \([V\text{-}NP_1\text{-}P\text{-}NP_2]\). For instance, in a sentence 5.4, the verb ‘inform’ subcategorizes a prepositional phrase in which the preposition is ‘of’.

(5.4) John informed the police of the danger.

To fill the argument position of the verb, the of-PP, \textit{i.e.}, ‘of the danger’ is obviously attached to the verb ‘inform’ as the verb expects an of-PP, and, in turn, the possibility of its attachment to the preceding noun, \textit{i.e.}, ‘police’ is ruled out. To use this information, we have introduced this property of verb as an attribute in
the dictionary. In other words, we enrich the dictionary by providing argument structure information of the lexical item. The lexical entry of the verb ‘inform’ looks like 5.5.

(5.5) \([\text{inform}] \{\text{`inform(icl>communicate)`}\}
(VRB, VOA, VOАCOMM, #_OF_AR2, #_OF_AR2_obj) \langle E, 0, 0 \rangle\)

In 5.5, the attribute \#_OF_AR2 in the attribute list indicates that the verb ‘inform’ takes an of-PP as its second argument, and the attribute \#_OF_AR2_obj implies that this of-PP is in obj relation with the verb. The lexical entries for nouns and adjectives are enriched in the similar fashion. On the same lines, the attributes for other prepositions are formed.

5.2.2 Determination of UNL Relations

Once the attachment ambiguity is resolved, the semantic relation is established between the two related words. In this section we will show how semantic relations are generated. Given a triplet \([X-P-N]\) where \(X\) is a noun or a verb and \(N\) is a noun phrase, the semantic relation between \(X\) and \(N\) depends on the functional meaning of the preposition-\(P\) and semantic properties of the head of the phrase (\(X\)) and the object of the preposition (\(N\)). Consider again the example in 5.6.

(5.6) John gave a flower to Mary.

Here, the noun ‘Mary’ is related to the verb ‘give’ and the relation between them is goal [16]. This is illustrated in figure 5.1. It is an idiosyncratic property of the word which selects the particular type of argument. Beth Levin’s work [2] is used extensively during the analysis and assigning appropriate UNL relations. It states that if the given verb falls in some semantic category or class of verbs which are syntactically and semantically similar in behavior then
they select similar arguments and semantic roles. For instance, a set of verbs termed as ‘give verbs’ have the capacity of assigning the role of goal to their to-prepositional phrase arguments. With the help of these features Enconverter forms the relation between the verb and the noun.

![Figure 5.1: The UNL graph for the sentence 5.6](image)

5.2.3 Design of the Rule Base to Handle Prepositional Phrases

Having the dictionary attributes in place, we shall now state the algorithm for all cases of prepositional phrase attachment constructs. We shall restrict our discussion only to the analysis rules required for resolving the attachment ambiguity.

In the frame \([V-NP_1-P-NP_2]\), provided the argument structure attributes for \(V\) and \(NP_1\), \(NP_2\) is attached to the verb \(V\) or noun \(NP_1\) in following four ways.

\[(5.7)\]

a. \(NP_2\) attaches to the verb \(V\), if verb specifies the attribute \(#_P\_AR_2\) in its lexical entry and \(NP_1\) does not specify \(#_AR_1\). Else, \(NP_2\) attaches to the noun \(NP_1\) in cases (b), (c) & (d).

b. The verb does not specify \(#_P\_AR_2\) and \(NP_1\) specifies \(#_P\_AR_1\).
The rules are added to the existing rule base so that they can work in association with other rules to make the complete analysis of a sentence. The new rules take into account the new attributes introduced into the dictionary. For example, the Rules r1 and r2 in 5.8 decide when to shift right leading to case b, c and d corresponding to the NP attachment. They are implemented for the for-prepositional phrases. The rules state to shift the cursor (windows) to the words on the right side of the sentence so as to achieve noun attachment.

(5.8) r1. \( R\{V,_{\text{FORAR2}}\}\{N,_{\text{FORAR1}}\}(\text{PRE,} _{\text{FOR}})P60; \)

r2. \( R\{V,^\text{,}_{\text{FORAR2}}\}\{N\}(\text{PRE,} _{\text{FOR}})P60; \)

The rule r3 in 5.9 creates an obj relation between V and N\(_1\) and in the next step relation is created between V and NP\(_2\). This rule leads to case (a), i.e., verb attachment. It states to create a relation between verb and its immediate object, i.e., NP\(_1\) and then move to the PP.

(5.9) r3.
\(<\{V,_{\text{FORAR2}},_{\text{FORAR2}}\text{obj}_2\}\{N,^\text{,}_{\text{FORAR1}}\text{obj}_1\}(\text{PRE,} _{\text{FOR}})P30; \)

The rule r4 in 5.10 creates a rsn (reason) relation between V and NP\(_2\) and then, deletes the node corresponding to NP\(_2\). It corresponds to the verb attachment.

(5.10) r4. \(<\{V,_{\text{FORAR2}},_{\text{FORAR2}}\text{rsn}_2\}\{N,\text{FORRES,PRERES}\text{rsn}_1\}(\text{PRE,} _{\text{FOR}})P25; \)
5.3 Evaluation Process

The experiment of generating UNL expressions of sentences with [V-NP₁-P-NP₂] frame was performed on the British National Corpus [3] and Wall Street Journal Corpus. The BNC corpus was chosen mainly because of its wide domain coverage. The only hindrance in using it is that the sentences are too long to be easily processed by the machine. Hence a word limit of 12-15 words per sentence was imposed on the test sentences. The steps in the evaluation are as follows:

a. The sentences with different patterns are extracted. Out of these, the sentences with phrasal verbs are filtered and removed.

b. These sentences are processed by the Enconverter to generate UNL expressions.

c. The correctness of the UNL expressions is manually ascertained. A correct UNL entails that attachment problem have been already solved.

The next two chapters will describe in detail the analysis of ‘to’ and ‘of’ prepositional phrases. Besides attachment problem, these prepositions participate in a few other linguistic phenomenon. These chapters will present the comprehensive analysis of these phenomenon.
Chapter 6

Processing of ‘of’

In this chapter we discuss of-prepositional phrase and special cases in its analysis. Besides prepositional phrase attachment it also has a unique challenge of its own which will be discussed in the following sections in detail. For example, noun phrases involving of-prepositional phrases are associative or partitive type, on which the semantic head depends. We discuss handling of such constructs in UNL framework.

6.1 Introduction

Handling of-PP constructs correctly is a must, for example, in machine translation. Given a sentence containing the frame [V-NP₁-of-NP₂], the problem is to determine the following:

(6.1) a. attachment site for of-PP₂ (whether the attachment site is V or NP₁)

b. headedness (if the attachment site is NP₁, then to detect the semantic head between NP₁ and NP₂)

c. semantic relation between the head (V, N₁ or N₂) and the tail (N₁ or N₂)
In order to resolve these issues, we have taken linguistic insights from Levin [2], Grimshaw [7] related to the preposition ‘of’. The linguistic representation of of-PP construction is presented in section 6.2. Implementation details to generate accurate UNL expressions are discussed in section 6.3, and evaluation result is given in section 6.4.

6.2 Distribution of of-NP Constructions in English

In this section we address the linguistic complexity involved in the of-PP construction, keeping in mind the problems mentioned in example 6.1. We observe that of-PP construction appears in various syntactic environments:

- [NP₁-of-NP₂-V]
- [V-NP₁-of-NP₂]
- [V-NP₁-of-NP₂-of-NP₃]
- [V-of-NP]
- [V-A-of-NP]

However, in this chapter, we shall restrict our discussion to the first three frames.

6.2.1 The Frame [NP₁-of-NP₂]

The research issues involved with the frame [NP₁-of-NP₂] are: detecting head and semantic relation between NP₁ and NP₂. Note that any lexical category phrase has a syntactic head and a semantic head. But a syntactic head always does not have to be identical to the semantic head. Accordingly three types of constructions are often discussed in linguistic literature: associative, partitive and kind construction.
**Associative Construction**

In an associative of-PP construction, *e.g.*, NP$_1$-of-NP$_2$, [of-NP$_2$] is an associative modifier of NP$_1$. For example, in sentence 6.2

(6.2) A donation of $50,000 could have been made to the charity.

The [of-NP$_2$] is interpreted as an argument of NP$_1$, *i.e.*, ‘*donation*’. Therefore, the lexical entry of the nouns, *viz.*, ‘*donation*’ has to treat this of-PP as a co-indexed argument rather than as an adjunct in its conceptual structure.

**Partitive Construction**

The partitive construction is that where NP$_1$, the syntactic head of noun phrase is not the semantic head of the phrase. It comprises of a determiner phrase (NP$_1$) and NP$_2$, as illustrated in example 6.3.

(6.3) a bundle of rags

The expression in example 6.3 refers to the quantification of rag. The syntactic head here is ‘*bundle*’ but the semantic head is ‘*rag*’. The status of the lexical element ‘of’ is partitive. One important observation in a partitive construction [NP$_1$-of-NP$_2$] is that the N$_2$ is the semantic head unlike the head of an associative construction [NP$_1$-of-NP$_2$] given in example 6.2. Table 6.1 presents a comprehensive distribution of the determiners/nouns in partitive constructions.

**Kind-Construction**

The kind-construction is attested in English in two different orders: (i) kind-initial and (ii) kind-final, as illustrated in example 6.4a and 6.4b.

(6.4)  a. a bird of that kind
Possible classes of NP₁ (syntactic heads) in a partitive construction | Examples
---|---
Whole and Fractional numbers | 1, 3, 98; one-third, three-fourth, ...
Aggregate Numbers | hundreds, thousands, millions, billions,...
Dozen words | dozen, ream, quire, gross, ...
Quantitative Determiners | either, neither, each, some, all, both, half, many, ...
Container words | can, bag, bottle, spoon, tin, ...
Collection words | group, bundle, bunch, herd, crowd, swarm, team, ...
Measure units | gram, Pound, gallon, inch, ton, ...
Indefinite amount | drop, pinch, dose, ...

Table 6.1: Different kind of NP₁ in partitive constructions

b. that kind of bird

This construction consists of a common noun (content noun) and a noun of a special class kind-words (viz., ‘kind’, ‘type’, ‘sort’, ‘variety’, ‘species’). In both kind-initial and kind-final constructions, the kind-NP is analyzed to be a modifier of the content noun, the semantic head of the construction.

6.2.2 The Frame [V-NP₁-of-NP₂]

The complexity of the frame [V-NP₁-of-NP₂] involves all the three research issues mentioned in section 6.1 above. Let us consider these examples.

(6.5) a. John [drained [the sink] [of the water]].

b. I [remind [the committee] [of [the suggestion]]].
c. A good listener [knows [the value of silence]].

In example 6.5, all the post-verbal complex NPs are in the frame [V-NP$_1$-of-NP$_2$]. The of-theme adjunct rule is applied to interpret the status of of-PP in the sentences like 6.5a. Therefore, it is attached to the verb ‘drain’ and not to the NP$_1$ ‘the sink’. In example 6.5b, the PP ‘of the suggestion’ can be interpreted as the theme of the verb ‘remind’, but it has the co-indexed argument status rather than as an adjunct. In example 6.5c of-NP$_2$ is not attached to the verb ‘know’; it is an associative modifier of NP$_1$, i.e., ‘the value’. The algorithm for the attachment cases is mentioned in the next section.

### 6.2.3 The Frame [V-NP$_1$-of-NP$_2$-of-NP$_3$]

The problem involved in the frame [V-NP$_1$-of-NP$_2$-of-NP$_3$] is little complex. This frame allows two of-PPs as the post-verbal elements, as illustrated in example 6.6.

(6.6) I [[informed [the author of the book]] of this comment].

[V-(NP$_1$-P$_1$-NP$_2$)-P$_2$-NP$_3$]

(6.7) He [[informed the registrar] of [his change of Allegiance]].

[V-NP$_1$-P$_1$-(NP$_2$-P$_2$-NP$_3$)]

The post-verbal elements in both the sentences in 6.6 & 6.7 are in the frame [V-NP$_1$-P$_1$-NP$_2$-P$_2$-NP$_3$]. In 6.6, a frame like this the possible attachment site for [P$_1$-NP$_2$] is either N$_1$ or V, and, the possible attachment site for [P$_2$-NP$_3$] is either N$_2$ or N$_1$ or V. To disambiguate the attachment site for [P$_1$-NP$_2$] and [P$_2$-NP$_3$], let us look at the conceptual structure of various lexical elements in 6.6. In 6.6 the complex NP, i.e., [NP$_1$-P$_1$-NP$_2$]: ‘the author of the book’
is considered as an argument of the verb, and, the PP, i.e., \([P_2-NP_3]\): ‘of this comment’ is considered as the second argument of the verb ‘inform’. Notice that the \([P_1-NP_2]\), i.e., ‘of the book’ is attached to \(N_1\), i.e., ‘author’. It is because of the fact that \(N_1\), i.e., ‘author’ demands an of-PP argument in its conceptual structure. At the same time the verb ‘inform’ also demands an of-PP argument; therefore the \([P_2-NP_3]\), i.e., ‘of this comment’ is attached to the verb.

The situation is little different in 6.7. The verb ‘inform’ subcategorizes an of-PP in its conceptual structure, but the \(N_1\), i.e., ‘registrar’ does not demand an of-PP argument nor an of-NP adjunct. Therefore, the PP, i.e., [of \([NP\) his change of Allegiance\)], is attached to the verb ‘inform’. Now the immediate requirement of the verb ‘inform’ in its conceptual structure is fulfilled. So the possible attachment site for \([P_2-NP_3]\), i.e., ‘of Allegiance’, is either \(N_1\) or \(N_2\). It is observed that \(N_1\), i.e., ‘the register’, does not require an of-PP argument nor an of-PP adjunct, and, on the other side \(N_2\) (i.e., ‘change’) requires an of-PP in its conceptual structure. Therefore, the attachment site for \([P_2-NP_3]\), i.e., ‘of Allegiance’ is \(N_2\), i.e., ‘change’.

6.2.4 Conclusions from the Linguistic Analysis

What we have discussed so far is summarized in this section. We have given a linguistic analysis for different of-PP constructions. In section 6.3 we shall present algorithm for Enconverter system of English. Table 6.2 exhibits the semantic heads pertaining to of-PP construction at a glance.
<table>
<thead>
<tr>
<th>Sub-Category of of-NP construction</th>
<th>Semantic Head</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associative</td>
<td>N₁/A</td>
<td>the value of silence aware of the situation</td>
</tr>
<tr>
<td>Partitive</td>
<td>N₂</td>
<td>two tons of rice</td>
</tr>
<tr>
<td>Kind-Construction</td>
<td>N₂(Kind-initial) or N₁(Kind-final)</td>
<td>an animal of that kind that kind of animal</td>
</tr>
<tr>
<td>of-adjunct</td>
<td>V</td>
<td>He cleared the desk of papers.</td>
</tr>
<tr>
<td>of-argument</td>
<td>V</td>
<td>The Police informed the parents of the accident.</td>
</tr>
</tbody>
</table>

Table 6.2: Semantic heads at a glance

6.3 Implementation of of-PP in UNL

In this section we shall develop the algorithm to process all the five frames in question by considering the linguistic insight developed so far in the previous section.

6.3.1 Augmenting the Lexical Entries with Argument Structure Information

As we have seen in last chapter we add argument structure information in terms of attributes in the dictionary. For instance, in a sentence 6.8, the verb ‘inform’ subcategorizes a PP in which the preposition is ‘of’.

(6.8) John informed the police of the danger.

The lexical entry of the verb ‘inform’ looks like 6.9.
In (6.9) the notation \#_{AR2} in the attribute list indicates that the verb inform takes an of-PP as its second argument, and the notation \#_{AR2, obj} implies that this of-PP is in obj relation with the verb.

6.3.2 Implementation to Handle the of-PPs

In this section we shall present the details of the rules applied for processing different of-PP constructs and we shall restrict our discussion to the most difficult cases.

The Frame \[V-NP_1-of-NP_2\] in UNL

In this section we shall illustrate the algorithm for the frame \[V-NP_1-of-NP_2\]. Figure 6.1 illustrates three attachment possibilities for NP_2.

To account for these three attachment possibilities, four different cases of V and N_1 combination are considered. The algorithm de-
The rules are added to the existing rule base so that they can work in association with other rules to make the complete analysis of a sentence.

### 6.3.3 The Strategy

What we have discussed so far for handling of five different frames in UNL is summarized in figure 6.2.

### 6.4 Evaluation

The experiment of generating UNL expressions from of-PP constructs was performed on the British National Corpus [3] and Wall Street Journal Corpus. The BNC corpus was chosen mainly because of its wide domain coverage. The only hindrance to using it is that the sentences are too long to be easily processed by machine. Hence
a word limit of 12-15 words per sentence was imposed on the test sentences. The steps in the evaluation are as follows.

a. Sentences with different patterns are extracted. Care is taken to isolate frames with phrasal verbs which are not in the scope of the current work.

b. These are processed by the Enconverter to generate UNL expressions.

c. The correctness of the UNL expressions is manually ascertained. A correct UNL entails that attachment and headedness problems have been already solved.

Many top level statistics of the system performance is given in table 6.4 and 6.5.

<table>
<thead>
<tr>
<th># [NP-of-NP]</th>
<th>= 1140</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of partitive cases</td>
<td>= 17.3% (197)</td>
</tr>
<tr>
<td>Average accuracy for partitives</td>
<td>= 92% (182)</td>
</tr>
</tbody>
</table>

Table 6.4: Statistics of Partitive Cases in BNC

| #Total [V-N₁-of-N₂] | = 1000 |
| #Verb attachment cases | = 9 |
| #Noun attachment cases | = 991 |

Table 6.5: BNC statistics for attachment cases

The errors are generally due to inaccuracies in lexical attributes and the interference of words extraneous to the frame. The relatively low accuracy of the [V-A-of-N] is attributed to the fact that in this frame the V is usually a be verb or a raising verb (‘seem’, ‘appear’ etc.) whose analysis into the UNL form is nontrivial even manually. The high accuracy in case of partitives is largely due to
Table 6.6: Accuracy of UNL generation for different frames of of-NPs (BNC and WSJ data)

<table>
<thead>
<tr>
<th>Frames</th>
<th>No. of sentences</th>
<th>No. of correct attachments and UNL</th>
<th>No. of incorrect cases</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-of-N</td>
<td>100</td>
<td>87</td>
<td>13</td>
<td>87%</td>
</tr>
<tr>
<td>N1-of-N2-V</td>
<td>100</td>
<td>92</td>
<td>8</td>
<td>92%</td>
</tr>
<tr>
<td>V-N1-of-N2</td>
<td>1000</td>
<td>886</td>
<td>114</td>
<td>88%</td>
</tr>
<tr>
<td>V-N1-of-N2</td>
<td>661</td>
<td>597</td>
<td>64</td>
<td>90%</td>
</tr>
<tr>
<td>V-A-of-N</td>
<td>300</td>
<td>237</td>
<td>63</td>
<td>79%</td>
</tr>
<tr>
<td>V-N1-of-N2</td>
<td>50</td>
<td>39</td>
<td>11</td>
<td>78%</td>
</tr>
</tbody>
</table>

the correctness and completeness of the lexical attributes resulting from an in-depth linguistic study.

To sum up, our system reports 86% accuracy on BNC data and 90% accuracy on WSJ data, where three important issues (i.e., attachment site, headedness and semantic relation) have been addressed. In comparison to this, Ratnaparkhi’s stated accuracy [20] of 99% for the preposition ‘of’ and 82% for all prepositions is achieved on WSJ data, where only the PP-attachment issue is addressed.
Figure 6.2: General strategy
Chapter 7

Processing of ‘to’

‘To’ is one of the most common lexemes in English (as found in 34% of the sentences in the Penn Tree Bank corpus [9]). This chapter describes the analysis of the sentences involving ‘to’. It poses problems at lexical, syntactic and semantic level, as we need to resolve whether the phrase which is fronted by ‘to’ is a prepositional phrase (PP) or an infinitival clause which is a kind of inflectional phrase (IP). In case of PP, we need to locate the attachment site for it and assign the semantic relation. In case of IP, we need to resolve the empty pronominal in addition to the attachment resolution — a task essential to make the semantic analysis complete. We achieve an accuracy of about 80% in the above tasks.

7.1 Introduction

The lexeme ‘to’ can form prepositional phrase or inflectional phrase. The prepositional phrase is formed by ‘to’ followed by a noun phrase, whereas, the ‘to’ followed by a base form of a verb constitutes an infinitival clause. This is illustrated by examples 7.1 and 7.2a, respectively.
Table 7.1: Heuristics to detect to-prepositional phrase

- ‘to’ is followed by a determiner
- ‘to’ is followed by a number
- ‘to’ is followed by a plural noun
- ‘to’ is followed by an adjective
- ‘to’ is followed by a proper noun
- ‘to’ is followed by a pronoun
- the matrix verb/noun specifies that it needs a to-preposition complement.
- ‘to’ is followed by a singular noun

(7.1) He gave the book [to Mary]_{PP}.

(7.2) a. He wants [to go.]_{IP}

b. He, wants [PRO_{i} to go].

Since most of the nouns in English can also be used as verbs, given a sentence involving ‘to’, it is difficult for the machine to detect the role of ‘to’. Besides, the prepositional phrase or infinitival clause has to be attached appropriately. For example, as in 7.1, ‘to Mary’ attaches to the verb ‘give’ and not to ‘book’. Consider the sentence 7.2a. Here, the infinitival clause has the implicit subject which is same as the subject in the main sentence, i.e., ‘he’. This issue of resolving the empty pronominal is also taken care of. This is essential, since we ultimately produce the meaning representation of the sentences called Universal Networking Language (UNL) graph.
7.2 Detecting the Role of ‘to’

Sentences containing ‘to’ were analyzed for their syntactic patterns, with the help of the corpora [3][1]. This analysis revealed the heuristics for the correct role detection of ‘to’ (see table 7.1 and 7.2).

7.3 Analysis of to-prepositional Phrase

Once the role of ‘to’ is determined, the next task is to find the correct attachment site for the to-prepositional phrase. In the frame [V-NP1-to-NP2], the to-prepositional phrase can attach to NP1 or to V. Consider sentences 7.3 and 7.4. In 7.3 the noun ‘Mary’ attaches to the verb ‘give’, whereas in (7.4) the noun ‘essay’ attaches to the noun ‘amendment’.

(7.3) John gave a flower to Mary.

(7.4) She made several minor amendments to her essay.

The algorithm [11] developed in case of analysis of ‘of’, uses the argument taking property of the verbs and the nouns called the subcategorization frame. It states that

“In the frame [V-NP1-to-NP2], attach NP2 to V or NP1
Table 7.3: Four cases of determining attachment site for to-prepositional phrase

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Action</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument of V</td>
<td>Argument of N1</td>
<td>N2 attaches to</td>
<td>UNL relation between V/N1 and N2</td>
</tr>
<tr>
<td>1</td>
<td>No TO</td>
<td>No TO</td>
<td>N1</td>
</tr>
<tr>
<td>2</td>
<td>No TO</td>
<td>TO</td>
<td>N1</td>
</tr>
<tr>
<td>3</td>
<td>TO</td>
<td>No TO</td>
<td>V</td>
</tr>
<tr>
<td>4</td>
<td>TO</td>
<td>TO</td>
<td>N1</td>
</tr>
</tbody>
</table>

Whether the verb or noun takes an argument is mentioned along with other information in the lexicon. Depending on the presence or absence of these attributes, four possibilities arise. Those are listed in the table 7.3.

### 7.4 Analysis of *to*-infinitival Clause

In this section, we deal with two problems, namely, attachment and PRO resolution concerning *to*-infinitival clause.
7.4.1 The Attachment Issue

In the frame \([V_1\text{-NP-}to-V_2]\), there exists a syntactic level to-infinitive clause attachment ambiguity. Consider sentences 7.5 and 7.6.

(7.5) A mild form of exercise will increase your ability [to relax].

(7.6) He ordered us [to finish the work].

In example 7.5, ‘to relax’ is attached to noun phrase ‘your ability’ and not to the verb ‘increase’, whereas in (7.6), ‘to finish the work’ is attached to the verb ‘order’. The nature of this ambiguity is exactly same as the prepositional phrase attachment ambiguity, if the frame \([V_1\text{-NP-}to-V_2]\) is considered. Thus, a similar strategy is used to solve the to-infinitive clause attachment ambiguity. Considering the fact that to-infinitival clause has more affinity to attach to a verb rather than to a nominal concept, in the first case the attachment preference is given to the \(V_1\) rather than to NP. The modifications are presented in table 7.4.

7.4.2 PRO in to-infinitivals

In the sentences with to-infinitival clauses, the subject of the embedded clause is sometimes not phonetically realized which is termed as PRO. Consider sentence 7.7.

(7.7) He promised me [to come for the party].

In this sentence, ‘to come for the party’ is an infinitival clause. But the subject of this infinitival clause is missing. The reading of 7.7 suggests that the subject is ‘he’, i.e., same as subject of the main sentence. This missing subject is indicated as PRO as shown in 7.8. In other words, it is said that this PRO is co-indexed with the subject of the matrix clause, i.e., ‘he’.
<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Action</th>
<th>V2 attaches to</th>
<th>UNL relation between V1/N and V2</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument of V1</td>
<td>Argument of N</td>
<td>Argument of V1</td>
<td>V2 attaches to</td>
<td>UNL relation between V1/N and V2</td>
<td>Examples</td>
</tr>
<tr>
<td>1 No TO</td>
<td>No TO</td>
<td>V1</td>
<td>obj (gather, battle)</td>
<td></td>
<td>They gathered\textsubscript{V1} a significant crowd to battle\textsubscript{V2} for their demand.</td>
</tr>
<tr>
<td>2 No TO</td>
<td>TO</td>
<td>N</td>
<td>obj (ability, relax)</td>
<td></td>
<td>A mild form of exercise will increase your ability\textsubscript{N} to relax\textsubscript{V2}.</td>
</tr>
<tr>
<td>3 TO</td>
<td>No TO</td>
<td>V1</td>
<td>pur (order, leave)</td>
<td></td>
<td>John ordered\textsubscript{V1} Mary to leave\textsubscript{V2}.</td>
</tr>
<tr>
<td>4 TO</td>
<td>TO</td>
<td>N</td>
<td>pur (effort, resolve)</td>
<td></td>
<td>He urged a bipartisan effort\textsubscript{N} to resolve\textsubscript{V2}.</td>
</tr>
</tbody>
</table>

Table 7.4: Four cases of to-infinitive clause attachment
(7.8) He promised me [PRO to come for the party].

Now, consider the sentence 7.9.

(7.9) He ordered us [to finish the work].

Here, it can be discerned that the missing subject of the embedded clause refers to ‘us’ which is also the object of the main sentence. As shown in 7.10.

(7.10) He ordered us [PRO to finish the work].

It can be observed from 7.7 to 7.10 that the PRO can be co-indexed with subject of the main sentence or object of the main sentence.

Thus, depending on the site of co-indexing, PRO is said to be subject-controlled or object-controlled [10]. In addition to this, the subject-controlled or object-controlled PRO is the idiosyncratic property of the verb of the main sentence. It is also noteworthy that [V-N-to-V] frame does not necessarily have a PRO. Consider the example 7.11. In this sentence the subject of the infinitival clause is ‘John’, i.e., the subject is overtly present for the to-infinitival clause and it is not the argument of the verb ‘want’ of the main sentence. This reading is shown in 7.11 with explicit bracketing. Since this type of behavior is not common, it is not considered in the analysis.

(7.11) I want John to do this work.

(7.12) I want [John to do this work].

Depending on the syntactic frame and the control property of PRO, verbs are classified into three types as seen in table 7.5. This property is also incorporated into the dictionary as a feature. In case of to-infinitival clause which attach to a noun, may have PRO as shown in example 7.13, PRO resolution in such cases is not addressed in this paper.
Verb type | PRO con- | Frame | Example
---|---|---|---
1 Promise | Subject | V-to-V | They, promised [PRO, to give a party].
2 Force | Object | V-N-to-V | They forced Mary [PRO, to come early].
3 Try | Subject | V-to-V | They, tried [PRO, to give a party].

Table 7.5: Three cases of *to*-infinitive clause attachment

(7.13) A mild exercise will increase your ability [to relax].

7.5 Dictionary

Enconverter, the analyzer machine, is driven by the dictionary. Each dictionary entry has a set of features associated with it. This set of features include both interpretable and uninterpretable features. The uninterpretable features are used for syntactic processing whereas the interpretable features are used for semantic interpretation. Some of those are not readily obtainable in the lexical resources available presently. We now discuss how to infer these features from various resources.

7.5.1 Acquisition of Subcategorization Frame Information

In English, the subcategorization property of lexical items is uninterpretable. The sub-categorization information for the words is gathered through two lexical resources, WordNet [6] and Oxford Ad-
vanced Learner’s Dictionary (OALD) [1]. OALD provides the partial syntactic frames against each word. These frames were mapped to UW dictionary as features indicating the preposition phrase and infinitival clause arguments. For example, consider the entry for the noun *effort* in OALD as shown in figure 7.1. The entry specifies that *effort* takes ‘*to do sth*’ type of frame which means the noun ‘*effort*’ takes *to*-infinitival phrase as its argument [7].

![Figure 7.1: Part of dictionary entry for ‘effort’](image)

The WordNet provides sentence frames against each sense of the verb, which also includes an example sentence particular to the verb. These sentences are actually prepared with the fixed number of template sentences. Sentence 7.14 shows one such template sentence for verb ‘*promise*’ with a placeholder for the verb.

(7.14) They _him to write the letter.

From all the sentence frames, those sentence frames which have a *to*-infinitive clause are identified. Then, for a given verb it can be easily ascertained if it takes *to*-infinitival clause as an argument or not. The frames also specify whether the argument is the first argument of the verb or the second.

### 7.5.2 Extracting PRO Information

We assume that the property of an empty pronominal PRO (i.e., whether the PRO is subject-controlled or object-controlled or ar-
bitrary) is interpretable in English. As discussed earlier, Subject-controlled PRO or Object-controlled PRO is the idiosyncratic property of the verb of the matrix clause. This section explains how a syntactically annotated corpus can be used to extract such semantic information. Penn Treebank [9] provides best syntactic parse trees along with providing consistent treatment to empty categories and subject marking for infinitival clauses and various other phenomena. A typical sentence involving to-infinitival clause from Penn Treebank is shown in figure 7.2. The *-1 in ‘NP-SBJ *-1’ is the trace of the subject of the to-infinitival clause and it is co-indexed with NP-SBJ-1 which is the noun ‘investors’.

Figure 7.2: The parse tree for ‘Investors continue to pour cash into money funds’ from Penn Treebank

The Algorithm to extract PRO information is given below. This feature is incorporated into the dictionary as an attribute.

For each sentence,

1. If the sentence contains ‘to’ under VP node (i.e., to-infinitive), find its subject NP, if it is not empty continue with next sentence, else,
   (a) Find the verb to which to-infinitival clause is attached.
   (b) Find its subject-NP.
2. If this subject-NP is referred by same index as that of the empty element in step 1, mark the verb as having ‘Subject Controlled PRO’.

3. Else, find the object-NP for the matrix verb found in step 1a. (This step will detect and consider passive appropriately)
   (a) If it is ‘referred’ by same index as that of the empty element in step 1, mark the verb as having ‘Object Controlled PRO’.

4. Find the root of the inflected verb with the help of WordNet.

5. End.

7.6 Assigning Semantic Relations

Once the attachment ambiguity is resolved, the semantic relation is established between the two related words. In this section we will show how semantic relations are generated.

7.6.1 UNL Generation for to-PP

Given a triplet [X-to-N] where X is a noun or a verb and N is a noun phrase, the semantic relation between X and N depends on the functional meaning of the preposition-to and semantic properties of the head of the phrase (X) and the object of the preposition (N). Consider again the example in case 3 of table 7.3 as (7.15).

\[(7.15)\] John gave a flower to Mary.

Here, the noun ‘Mary’ is related to the verb ‘give’ and the relation between them is goal [16]. This is illustrated in figure 7.3. It is an idiosyncratic property of the word which selects the particular
type of argument. Beth Levin’s work [2] is used extensively during
the analysis and assigning appropriate UNL relations. It states that
if the given verb falls in some semantic category or class of verbs
which are syntactically and semantically similar in behavior then
they select similar arguments and semantic roles. For instance, a set
of verbs termed as ‘give verbs’ have the capacity of assigning the role
of goal to their to-prepositional phrase arguments. These relations
are encoded in the dictionary as interpretable features. With the
help of these features Enconverter forms the relation between the
verb and the noun.

![Figure 7.3: The UNL graph for the sentence 7.15](image)

**7.6.2 UNL Generation for to-infinitival Clause**

The UNL has the notion of representing compound concepts, *i.e.*, a
concept composed of concepts and relations (compound UW ). Since
the to-infinitive clause is conceptualized as IP (inflectional phrase)
in X-bar schema, it is represented as a compound UW in UNL. The
PRO is realized as UW which is referred doubly, *i.e.*, in the main
sentence and in the clause. Consider the example 7.16.

(7.16) They promised Mary [PRO \_ to give a party].

The UNL representation for 7.16 is given in figure 7.4. The UW
for ‘they’ is related to ‘promise’ as well as to UW for ‘give’; *i.e.*,
The UW for ‘they’ is co-referred by UW for ‘promise’ as well as ‘give’. This co-referencing is reflected in the UNL in a way that the UW ‘they’ is used in relations involving both give(icl>do) and promise(icl>do). The to-infinitival clause is represented as single concept, indicated by large eclipse with label ‘:01’ and then related to promise(icl>do) through obj(object) relation. The processing of generating UNL relations for the triplet [X-to-V] where X can be a noun or a verb is similar to to-prepositional phrases.

![Figure 7.4: The proposed UNL graph for the sentence 7.16](image)

### 7.7 Strategy

Figure 7.5 shows the sequence of steps involved in the processing of the sentence involving the lexeme-to. Once the part of speech is identified the processing of both the prepositional phrase and infinitival phrase is done separately.

### 7.8 Evaluation

This section describes the creation of test-data, preparation of dictionary entries, the results and the error analysis.
7.8.1 Preparation of Test Sentences

The testing is performed on the sentences obtained from Penn treebank corpus, Oxford Advanced Learner’s Dictionary (OALD). All compound sentences, complex sentences except those involving to-infinitive clauses are not considered. Only declarative sentences are selected. All other sentences like interrogative, exclamatory sentences were filtered out.
7.8.2 Creation of Dictionary Entries

The dictionary was prepared for all the sentences selected as above. Both the interpretable and uninterpretable features are added to the dictionary as discussed in earlier section on dictionary. This dictionary is appended to the function word dictionary to form a complete dictionary.

7.8.3 Testing

With the help of the dictionary and rules, Enconverter convert the sentences to the UNL. The output is checked to see if the correct role of to is detected by the system. Also, we observed the correctness of UNL which entails the correct attachment. The result for 200 sentences tested is summarized in the table 7.6.

7.8.4 Error Analysis

The errors are mainly because of the inconsistent attributes in the dictionary and because of the inadequate rules. The following examples illustrate the various factors responsible for errors. Consider sentence 7.17a, the noun ‘refund’ is not the object of the matrix verb ‘order’ but it is the subject of infinitive clause as given in 7.17b. The

<table>
<thead>
<tr>
<th>Number of sentences</th>
<th>Preposition sense</th>
<th>Infinitive sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>correct sense of to is detected</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>correct attachment/UNL</td>
<td>84</td>
<td>77</td>
</tr>
</tbody>
</table>

Table 7.6: The results from testing of sentences with lexeme ‘to’
system treats it as an object of verb ‘order’ which leads to wrong UNL generation.

(7.17)  
  a. Judge ordered the refunds to begin in May.
  b. Judge ordered [the refunds to begin in May.]

In example 7.18a, the PRO (referred to ‘the industry’ as shown in example 7.18b) should relate to verb ‘continue’ with aoj relation. But instead, the system produces the default <relation> relation between the two, because of the lack of the rules to produce aoj relation in infinitival clause. This leads to wrong UNL generation.

(7.18)  
  a. Government helped the industry to continue.
  b. Government helped [the industry]i [PROi to continue].

(7.19)  
  a. They signed an agreement to acquire additional revenue.
  b. They signed [an agreement [PROx to acquire additional revenue.]]

In example 7.19a, there is an error in the attachment. Ideally, the infinitival clause should attach to noun ‘agreement’ and not to the verb ‘sign’. But since the noun ‘agreement’ does not specify a to-infinitival clause as an argument, the to-infinitival clause gets attached to verb ‘sign’ and not to the noun ‘agreement’.
Chapter 8

Analysis of Temporal Prepositional Phrases

In this chapter we show how an in-depth analysis of the properties of prepositions improves the efficacy of the process of analysis of English language. Our focus is on temporal prepositions of English and the UNL case relations they produce. The chapter ends with the section discussing the results obtained in this analysis.

8.1 Structure of Temporal Prepositions

Prepositions are syntactic connecting words. On the basis of their surface value, it is very difficult to categorize prepositions. But they can be classified into those of place, time, position, instrument, manner etc. In this section, we present the conceptual structure of temporal prepositions.

8.1.1 Temporal Field

According to Jackendoff, the semantic parts of speech [13], include Thing, Event, State, Action, Place, Path, Property and Amount.
Prepositions of time are similar in behavior as that of spatial expressions and that temporal preposition phrases are attached to sentences in the same way as prepositional phrases of location. In other words, the temporal expressions define a one-dimensional time-line. Generally, the event and state nouns specifies the time information. Let us consider the syntactic and conceptual structure of a prepositional phrase in 8.1 that refers to time.

(8.1) At noon

- Syntactic structure: \([PP_{at} \ [NP_{noon}]]\)
- Conceptual structure: \([Place \ ATemp \ [Time_{noon}]\]

According to conceptual structure, the temporal property of ‘at’ and ‘noon’ matches or in other words the temporal sense of ‘at’ selects temporal sense of ‘noon’. The reading of this temporal preposition phrase is illustrated in a sentence like 8.2.

(8.2) He came to office at noon.

The temporal preposition phrase at noon in 8.2 is not subcategorized by the head ‘come’. On the other hand, the preposition phrase ‘at noon’ is an adjunct of verb ‘come’. That is, it is attached to verb in the parse tree. It can be said that the verb ‘come’ attaches to an event indicated by the preposition phrase as shown in 8.3.

(8.3) \([Event \ [Place \ ATemp \ [Time_{noon}]]\]

### 8.1.2 Temporal Prepositions

We have identified twenty-one different prepositions which can indicate temporality. The following is the list of all the prepositions of temporal type.
“at, in, on, after, before, beyond, by, into, since, from, during, for, till, until, to, inside, over, through, between, of, within”

Each of these prepositions takes a noun possessing time property or event property, as its complement in an X-bar fashion to denote temporality. Consider the prepositional phrases for ‘at’, ‘in’ and ‘on’, illustrated in sentences 8.4–8.6.

(8.4) The conference opens at(*on) noon.

(8.5) We are going away in(*at) the Summer.

(8.6) We shall meet on(*in) Monday.

In each case, the referred meaning is temporal. But it is not possible to construct phrases like ‘*on noon’, ‘*at the summer’, ‘*in Monday’. One of the reasons for this could be selectional restriction. In sentence 8.4, the preposition ‘at’ selects a precise time argument. Accordingly, the conceptual structure of the temporal preposition ‘at’ allows a noun that indicates precise time. In sentence 8.5, ‘on’ selects days and date argument, and in sentence 8.6, ‘in’ selects long period argument. In the conceptual structure of [PP in the summer] and [PP on Monday], the selectional restriction on the arguments appear as the conceptual information long period, day and date, respectively. In sentences 8.4–8.6, each prepositional phrase refers to a segment/point in the time-line where the event occurs. Further, a prepositional phrase also refers to a start-time or end-time, interval-time or duration of an event based on the conceptual structure of the preposition. For instance, the preposition ‘since’ selects a noun argument that can be either a time or an event, as illustrated in sentence 8.7.

(8.7) We have been here since 1994/Christmas.
<table>
<thead>
<tr>
<th>UNL Relation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tim</td>
<td>A time of an event</td>
</tr>
<tr>
<td>tmf</td>
<td>A time when an event starts</td>
</tr>
<tr>
<td>tmt</td>
<td>A time when an event ends</td>
</tr>
<tr>
<td>dur</td>
<td>A period of time during which an event occurs</td>
</tr>
<tr>
<td>seq</td>
<td>Two events in series</td>
</tr>
<tr>
<td>obj</td>
<td>A thing in focus that is directly affected by an event</td>
</tr>
</tbody>
</table>

Table 8.1: Temporal relations in UNL

Conceptual structure: \[ \text{Place} \ 	ext{SINCE} \text{Temp} \ 	ext{Time} \ \text{(TIME/EVENT)} \]

In next section, we present how such conceptual information of temporal prepositions is expressed in UNL using a sophisticated set of attributes.

8.2 From Properties to Relations

The culmination of the detailed linguistic analysis of temporal prepositions is a bridge between the English preposition repository and UNL relation inventory.

8.2.1 UNL Temporal Relations Inventory

Five relations, viz., tim, tmt, tmf, dur, and seq, are provided to handle time expressions in UNL (see table 8.1) [16]. These relations along with the obj relation are used to represent the meaning of temporal preposition phrases.
8.2.2 The Mapping between Temporal Preposition Phrases and UNL Relations

Table 8.2 depicts the bridge between the prepositions and the UNL relations. The first two columns are self explanatory; the third column shows the attributes that are necessitated by the properties of the preposition in the first column.

The meanings of the attribute labels in the third column are explained with examples in table 8.3 in the next section. For example, the third column for ‘in’ in table 8.2 means the preposition can be used in 3 situations, e.g., (a) in February, (b) in 2004 and (c) in the morning.

8.2.3 Lexical Entry of Prepositions

Prepositions have different semantic roles depending upon the arguments they select and the sites they get attached to. Each preposition has a single entry in the lexicon. No semantic attribute is placed in the lexical entry. For instance, the lexical entry for the preposition ‘since’ is:

\[(8.8) [\text{since}] \langle \text{\textit{since}}, \#\text{SINCE} \rangle \langle E, 0, 0 \rangle;\]

Every preposition is represented as \#\langle preposition \rangle in the attribute list. This is used to fire the correct rule.

8.2.4 Temporal Attributes

Processing of temporal prepositional phrases by the Enconverter needs both static \textit{i.e.}, lexical attributes and dynamic \textit{i.e.}, run-time attributes. For example, the processing of the phrase ‘on 3rd May’ will disambiguate ‘3rd’ — which can denote an ordinal number — by invoking a run time attribute \textit{DAY} prompted by the lexical attribute
<table>
<thead>
<tr>
<th>Preposition</th>
<th>UNL Relations</th>
<th>Attributes of the Arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td>at</td>
<td>tim</td>
<td>[TIME,TIM_TOKEN] [N, TIME]</td>
</tr>
<tr>
<td>in</td>
<td>tim</td>
<td>[N, TIME, MONTH] [N, TIME, YEAR] [N, POOF_DAY]</td>
</tr>
<tr>
<td>on</td>
<td>tim</td>
<td>[N, TIME, DAY]</td>
</tr>
<tr>
<td>after</td>
<td>tim-obj</td>
<td>[TIME, TIM_TOKEN] [N, TIME] [N, EVENT]</td>
</tr>
<tr>
<td>before</td>
<td>tim-obj</td>
<td>[TIME, TIM_TOKEN] [N, TIME] [N, EVENT]</td>
</tr>
<tr>
<td>beyond</td>
<td>tim-obj</td>
<td>[TIME, TIM_TOKEN] [N, TIME]</td>
</tr>
<tr>
<td>by</td>
<td>tmt-obj</td>
<td>[TIME, TIM_TOKEN] [N, TIME, DAY] [N, TIME, MONTH] [N, TIME, YEAR]</td>
</tr>
<tr>
<td>since</td>
<td>tmt</td>
<td>[TIME, TIM_TOKEN] [N, TIME, DAY] [N, TIME, MONTH] [N, TIME, YEAR] [N, EVENT]</td>
</tr>
<tr>
<td>from</td>
<td>tmf-obj</td>
<td>[TIME, TIM_TOKEN] [N, TIME, DAY] [N, TIME, MONTH] [N, TIME, YEAR] [N, EVENT]</td>
</tr>
<tr>
<td>during</td>
<td>dur</td>
<td>[TIME, TIM_TOKEN] [N, TIME, UNIT] [N, EVENT]</td>
</tr>
<tr>
<td>between</td>
<td>dur-obj</td>
<td>[N, EVENT]</td>
</tr>
<tr>
<td>through</td>
<td>dur-obj</td>
<td>[N, TIME, UNIT] [N, EVENT]</td>
</tr>
<tr>
<td>for</td>
<td>dur-obj</td>
<td>[N, TIME, UNIT]</td>
</tr>
<tr>
<td>over</td>
<td>dur-obj</td>
<td>[N, TIME, UNIT]</td>
</tr>
<tr>
<td>till</td>
<td>tmt</td>
<td>[TIME, TIM_TOKEN] [N, TIME, DAY] [N, TIME, MONTH] [N, TIME, YEAR] [N, EVENT]</td>
</tr>
<tr>
<td>until</td>
<td>tmt</td>
<td>[TIME, TIM_TOKEN] [N, TIME, DAY] [N, TIME, MONTH] [N, TIME, YEAR] [N, EVENT]</td>
</tr>
<tr>
<td>inside</td>
<td>dur-obj</td>
<td>[N, TIME, UNIT]</td>
</tr>
<tr>
<td>within</td>
<td>dur-obj</td>
<td>[N, TIME, UNIT]</td>
</tr>
<tr>
<td>from(-to)</td>
<td>tmf(-tmt)</td>
<td>[TIME, TIM_TOKEN] [N, TIME] [N, EVENT]</td>
</tr>
<tr>
<td>between(-and)</td>
<td>dur-tmf-tmt</td>
<td>[TIME, TIM_TOKEN, ANDRES] [N, TIME, ANDRES] [N, EVENT, ANDRES]</td>
</tr>
<tr>
<td>into</td>
<td>dur-obj</td>
<td>[N, TIME]</td>
</tr>
<tr>
<td>to</td>
<td>tim</td>
<td>[TIME, TIM_TOKEN] [N, TIME]</td>
</tr>
<tr>
<td>of</td>
<td>tim</td>
<td>[N, TIME]</td>
</tr>
</tbody>
</table>

Table 8.2: Temporal Relations and Attributes
MONTH of ‘May’. Dates in all formats are recognized in our system - which is an instance of the named entity recognition problem.

A time-word is given a set of attributes in its lexical entry. Table 8.3 shows a representative set of such attributes. The asterisk marked attributes are used as dynamic attributes also.

### 8.3 Formulation of Analysis Rules

We explain the creation of analysis rules for a few representative situations.

#### 8.3.1 General Strategy

a. **Basic Case with the frame [V P N]**: In this case the preposition in question translates directly into one of the temporal

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>Time indicating words</td>
<td>Today</td>
</tr>
<tr>
<td>UNIT</td>
<td>Unit of time-measurement</td>
<td>Day, second</td>
</tr>
<tr>
<td>EVENT</td>
<td>Events</td>
<td>Morning, Dinner</td>
</tr>
<tr>
<td>TIM_TOKEN</td>
<td>Indicates time</td>
<td>am, pm</td>
</tr>
<tr>
<td>SECOND*</td>
<td>Second</td>
<td>12, 15</td>
</tr>
<tr>
<td>MINUTE*</td>
<td>Minute</td>
<td>30, 45</td>
</tr>
<tr>
<td>HOUR*</td>
<td>Hour</td>
<td>6, 8</td>
</tr>
<tr>
<td>DAY*</td>
<td>Any day</td>
<td>Sunday; 23</td>
</tr>
<tr>
<td>POFDAY</td>
<td>Any event which is part of day</td>
<td>Dinner, Sunset</td>
</tr>
<tr>
<td>WEEK</td>
<td>Any week</td>
<td>2nd week</td>
</tr>
<tr>
<td>MONTH*</td>
<td>Any month</td>
<td>May, December; 11, 12</td>
</tr>
<tr>
<td>YEAR*</td>
<td>Any year</td>
<td>1996, 2004</td>
</tr>
<tr>
<td>ZDIM</td>
<td>Zero-dimensional word</td>
<td>Start, End, Middle</td>
</tr>
</tbody>
</table>

Table 8.3: Temporal Attributes
relation in UNL. The analyzer machine is directed by a set of rules to position its left analysis window on P and its right Analysis Window on N (vide figure 8.1). Now the lexical property of N and P are read off from their dictionary entries. Then the properties of preposition P are copied to the node of N and P is deleted. This brings V and N adjacent to each other. The analysis heads are positioned on these nodes and the \textit{tim} relation is produced between them.

(8.9) come at noon
  - Delete ‘at’
  - Produce \textit{tim}(come, noon)

b. [V P N] producing Time-Relation + \textit{obj}: In this case the function of the preposition does not match any of the temporal UNL relation. Thus to preserve the meaning of the particular function of the preposition, it is retained as UW and related with its syntactic object with \textit{obj} relation in UNL. Analysis windows are placed on P and N. The properties of P are transferred to N. The relation \textit{obj} is generated between P and N. This time N gets deleted. After deletion an appropriate temporal relation — dictated by the properties of P — is generated between V and P, as shown in 8.10.

(8.10) come before noon
c. Frame with preposition phrase attachment question: \([V N_1 P N_2] \) (V attachment) Produce obj between V and N_1, deleting N_1. Subsequent steps are same as in (a).

(8.11) attend meetings at noon
- Produce obj(attend, meeting)
- Delete ‘meeting’
- Produce tim(attend, noon) as case (a)

d. Frame with preposition phrase attachment question: \([V N_1 P N_2] \) (N_1 attachment) Deal with the N_1 P N_2 frame as in case (a). Then produce a UNL relation from this frame. Finally link the N_1 to V with the appropriate relation.

(8.12) attend the meeting of January
- Produce tim(meeting, January)
- Delete ‘January’.
- Then produce obj(attend, meeting)

This general strategy\(^1\) is pictorially depicted in figure 8.2.

8.4 Evaluation

The experiment of generating UNL expressions from sentences with temporal prepositional phrases was performed on the Wall Street Journal (WSJ) corpora, Brown Corpus [4] and sentences from Oxford Advanced Learner’s Dictionary of English Language [1]. It

\(^1\)The expression rel(x,y) refers to UNL relation between UWs x and y.
should be noted that the correctness of UNL generation entails correct prepositional phrase attachment. Various top level statistics of the system performance is given in table 8.4, followed by preposition accuracy distribution in figure 8.3.

It is seen from the results that the average accuracy is fairly high (about 84%). The observations also support the intuition that pure
temporal prepositions like until should give very high accuracy (vide figure 8.3).

The inaccuracies result from errors in the corpus (missing nouns, hyphenated words, spelling mistakes, nonsense words etc.) and also lexicographers’ mistakes in creating accurate dictionary entries, especially the grammatico-semantic attributes.
Chapter 9

Conclusion and Future Work

9.1 Conclusion

The current work aims to improve the accuracy of English analysis system which is part of UNL based English to Hindi machine translation system. This project concentrated on attachment problem of prepositional phrase and semantic relation generation in English. Our approach is knowledge based which uses a large repository of rules covering various language phenomenon for accomplishing the analysis task. This is also accompanied by the feature rich concept dictionary.

We attempted to capture the roles of all English prepositions in different possible contexts. To this effect, we used corpus extensively, to study the syntactic behavioral patterns of English sentences.

In the next experiment, we did the detailed analysis of of-PP constructs and the to-PP constructs. The of-PP has problems of its own such as identifying the semantic head in the noun phrase which
has of-PP. Besides, it has attachment problem and semantic relation determination problem. During the work on of-PP, we came up with the general strategy to solve prepositional phrase attachment which is heavily dependent on argument structure information. The work on ‘to’ involves analysis of to-infinitival and to-prepositional phrases. It also does the disambiguation of the role of lexeme ‘to’ (head of infinitival clause versus prepositional sense). The analysis of infinitival clause involves clause detection, attachment, PRO resolution etc. This is based on syntactic and semantic features available in the lexicon.

We have also attempted to extract such features through the non-structured resources. The work leads to enrichment of the lexicon. It, in turn, eased the creation of rule bases for knowledge extraction into the UNL form. The results of experiments show the usefulness of such analysis for machine translation. Also, the automatic acquisition of syntacto-semantic features such as subcategorization frame, selectional preference, thematic relation, is one of the major output of the project. Initially, we worked on classifying different aspects of prepositions. This leads us to study and classify all the prepositions which exhibit temporal roles.

In all, this work is a major step in improving the quality and accuracy of UNL produced from English text.

9.2 Future Work

Though the detailed work done so far involve only the lexeme ‘to’ and ‘of’, this approach will be applicable to all prepositions with no or minor changes.

The approach used for handling prepositional phrase attachment, too, can be extended to solve the clause attachment problem. We
could use the fact that nouns and verbs also specify if they have clause as its argument.

Two interesting research directions naturally arise from the work on PP attachment. It may be easily discerned that the lexical enrichments are nothing but selectional preference, subcategorization, argument structure and such information of the constituents of the VP. A verb knowledge base [19] and various other ontologies lends the possibility of automatically gleaning the syntacto-semantic attributes. Also the same methodology can be extended towards solving other constructs such as clause attachments. The other research direction is the use of stochastic methods in combination to arrive automatically at the rules for the Enconversion process. This will need a UNL corpus of a sizable number of sentences.
Bibliography


 Parsing. European Union Working Conference on Sharing Capability in Localization and Human Language Technologies (SCALLA04), Kathmandu, Nepal


