VERB TRANSFER FOR ENGLISH TO URDU MACHINE TRANSLATION
(Using Lexical Functional Grammar (LFG))

MS Thesis

Submitted in Partial Fulfillment
of the Requirements for the
Degree of

Master of Science (Computer Science)

at the

National University of Computer & Emerging Sciences

by

Nayyara Karamat

December, 2006

Approved:

_____
Head
(Department of Computer Science)

_____20 ____

Approved by Committee Members:

Advisor

_____

Dr. Sarmad Hussain
Professor
National University of Computer & Emerging Sciences

Other Member:

_____

Dr. Miriam Butt
Professor
Universitaet Konstanz, Germany

TABLE OF CONTENTS

1 Introduction

The demand for language translation has greatly increased in recent times due to increasing cross-regional communication and the need for information exchange. Most material needs to be translated, including scientific and technical documentation, instruction manuals, legal documents, textbooks, publicity leaflets, newspaper reports etc. Some of this work is challenging and difficult but mostly it is tedious and repetitive and requires consistency and accuracy. It is becoming difficult for professional translators to meet the increasing demands of translation. In such a situation the assistance of computers can be used as a substitute (Hutchins and Somers 1992).

The main difficulty in automated translation of one natural language to another is varied structures and lexical choices for the same concept in different languages. Syntactic and semantic analysis is performed to reach a logical form of the language to be translated. The ultimate aim is to define a logical form that can represent the meaning of the text independent of any language. This level of representation would be ideal but is difficult to achieve. It is so difficult for analysis of any language to reach such an abstraction that to bridge the gap, some transfer mechanism is required.

The aim of this thesis is to look into translation issues raised by the transfer of verbs in English to Urdu machine translation.

First, in the background section of this thesis, the basic theory for machine translation systems, Lexical Functional Grammar (LFG) and grammatical analysis of verbs and translation problems is presented. Then the problem statement is defined which is followed by the methodology. The results of the study are then presented.

2 Background

In this section the necessary background information required to understand the problem statement will be provided. Section 2.1 gives a brief introduction to machine translation and different architectures of machine translation systems. Section 2.2 describes the basic notion of LFG. Section 2.3 gives an overview of LFG analyses for verbs in different languages. Section 2.4 describes some problems which are faced during translation from one language to another.

2.1 Machine Translation

The term Machine Translation (MT) can be defined as "translation from one natural language (source language (SL)) to another language (target language (TL)) using computerized systems, with or without human assistance" (Hutchins and Somers 1992, pg. 3).

Machine translation systems can be divided in two generations. First generation systems are known as *direct systems*. In such systems, translation is done word by word or phrase by phrase. In such systems very minimal linguistic analysis of input text is conducted (Hutchins and Somers 1992). This architecture is still being extensively used in commercial MT systems. The main idea behind direct systems is to analyze the input text to the extent that some transformational rules can be applied. This analysis could be parts of speech of words or some phrasal level information. Then using a bilingual dictionary, source language words are replaced with target language words and some rearrangement rules are used to modify the word order according to the target language (Arnold et al. 1993).

This architecture is very robust because it does not fail on any erroneous or ungrammatical input. Since the analysis level is very shallow and the system contains very limited grammatical information, it hardly considers anything ungrammatical. In the worst case if the rule does not apply to the input, the input is passed on without any

alteration as output. This kind of system is hard to extend because all the rules are written in one direction and are language specific. To make another language pair work, all the rules have to be re-written. Since the system does not perform very deep analysis, its time complexity is low. These systems work very well for closely related languages but are not suitable for modeling languages with diverse syntactic nature. Since the system does not explicitly contain the grammatical rules of the target language, there is a chance that the output will not be grammatical but it will be similar to the target language (Arnold et al. 1993).

Owing to the fact that linguistic information helps an MT system to produce better quality target language translation, with the advance of computing technology, MT researchers started to develop methods to capture and process the linguistics of sentences. This was when the era of second generation MT systems started. Second generation machine translation systems are called indirect systems. In such systems the source language structure is analyzed and text is transformed into a logical form. The target language translation is then generated from the logical form of the text (Hutchins and Somers 1992). The transition from direct systems to indirect systems is illustrated in Figure 2.1, taken from (Hutchins and Somers 1992, pg. 107).

SYSTRAN is one of the most well-known direct systems. It is described in Hutchins and Somers (1992) and Wilks (1992).

Indirect systems can be further divided into interlingua and transfer based systems.

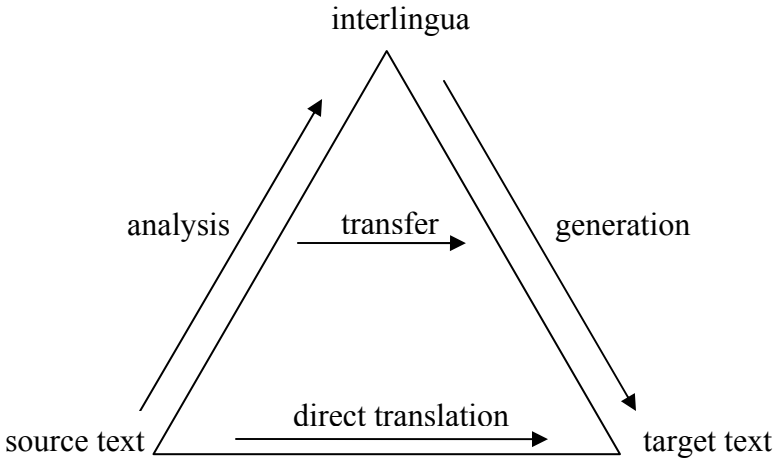

Figure 2.1: Transfer and interlingua 'pyramid' diagram

As shown in Figures 2.2a and 2.2b, the structures of these systems are fairly similar.

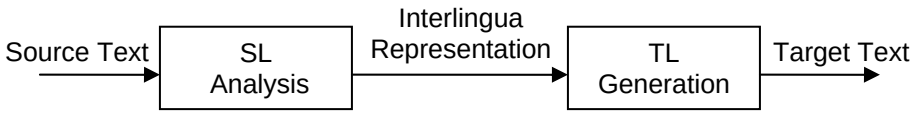

Figure 2.2a: Interlingua Based System

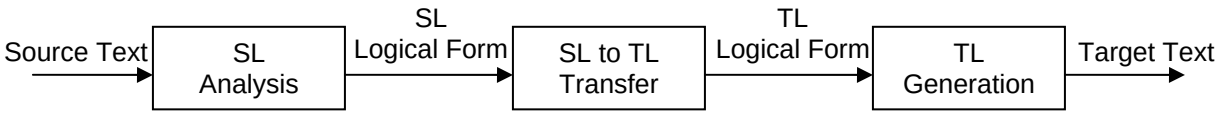

Figure 2.2b: Transfer Based System

In the transfer method, the source language is analyzed to an abstract level. Then, through a transfer module, this abstract form is converted to the corresponding abstract form in the target language through which the target translation text is generated.

The module '*SL Analysis*' captures the required linguistic information about the source language sentences to aid the translation. '*SL to TL Transfer*' module transfers the representation generated by '*SL Analysis*' to a target language representation. The module '*TL Generation*' generates the translation text using this logical representation. Such a system requires independent grammars for the source and target languages. Moreover it

requires a comparative grammar or transfer roles to relate source structures to target structures.

It is difficult to handle ungrammatical input using this approach. Since the system assumes full grammatical knowledge it does not allow ungrammatical sentences to be parsed, thus reducing the output of the system. This kind of system is easy to extend because to add a new language, grammar and transfer rules for the new language need to be written but the grammar of the other language is reusable. Such systems are theoretically reversible. The same grammars can be used in the reversed system. Practically there are problems in reversing the system because some transfer rules which are correct in one direction may not be correct in the other direction. The system has the explicit grammar of the target language, which ensures grammatical output (Arnold et al. 1993).

Examples of transfer systems include ARIANE (Vauquois and Boitet 1985), SUSY (Maas 1987), MU (the Japanese National Project) (Nagao et al. 1986), METAL (Slocum et al. 1987; Bennett and Slocum 1988), TAUM-AVIATION (Isabelle 1987), ETAP-2 (Apresian et al. 1992), LMT (McCord 1989), EUROTRA (Arnold 1986; Arnold and des Tombe 1987; Copeland et al. 1991a,b), CAT-2 (Sharp 1988), MIMO (Arnold and Sadler 1990), MIMO-2 (van Noord et al. 1990) and ELU (Estival et al. 1990).

The Interlingua approach involves the use of an intermediate language (i.e. an Interlingua) for the transfer, with the source language text translated to the Interlingua and the Interlingua translated to the target language text. As suggested by Hutchins and Somers (1992), an Interlingua is an intermediate 'meaning' representation and this representation:

"*includes all information necessary for the generation of the target text without 'looking back' to the original text. The representation is thus a projection from the source text and at the same time acts as the basis for the generation of the target text; it is an abstract representation of the target text as well as a representation of the source text.*" (Hutchins and Somers 1992, p. 73)

Interlingua appears to be an attractive approach for machine translation due to several reasons. Firstly, from a theoretical point of view it is very interesting to establish a representation which is independent of language. Secondly, Interlingua systems are more easily extendable because only analysis and generation modules are required to add a new language and no language specific transfer information is needed. But it is difficult to define such a language independent representation even for closely related languages (Arnold et al. 1993).

An attempt to define an Interlingua to represent the language in the form of a semantic relation is The Universal Networking Language (UNL) project. This project was initiated by the University of United Nations based in Tokyo in 1996. An utterance is represented as a hyper-graph in UNL. Normal nodes in the graph bear Universal Words (UWs) with semantic attributes and arcs bear semantic relations (deep cases, such as agt, obj, goal, etc.). UNL representation is being built in many languages including Arabic, Chinese, French, German, Hindi, Indonesian, Italian, Japanese, Mongolian, Portuguese, Russian, and Spanish.

Some other Interlingua systems are Rosetta (Landsbergen 1987b,a), KBMT (Goodman 1989; Goodman and Nirenburg 1991). (Arnold et al. 1993).

There are new emerging approaches to MT known as the empirical approaches. They apply statistical or pattern matching techniques for MT. These techniques are called empirical since the knowledge for translation is derived empirically by examining text instead of linguistic rules. There are two such approaches, the ‘example’ or ‘analogy’ based approach, and the ‘statistical’ approach (Arnold et al. 1993).

In the ‘example-based’ approach, translation is done by matching the given text with stored example translations. The basic idea is to collect a bilingual corpus of translation pairs and then use a best match algorithm to find the closest example to the source phrase to be translated. This gives a translation template, which can then be filled in by a word-for-word translation. A limitation of this technique is that it requires a large bilingual aligned corpus. But these examples can also be built incrementally, increasing the quality

of translation. Such systems are efficient because they need not to go through complex grammars to analyze the text, but if many examples match the input text then finding the best match can be a complex task. A pure example based system will include no linguistic knowledge but addition of some linguistic knowledge can improve the system by increasing its capability of dealing with more patterns concisely as one can specify categories instead of raw words (Arnold et al. 1993).

The second approach, the 'statistical approach', uses probabilistic analysis in MT as the name suggests. This term sometimes refers to the use of probability based techniques in parts of the MT task like word sense disambiguation or structural disambiguation. The other use of this term refers to a pure statistical machine translation system which uses probabilistic models to determine the correct translation of input text. In this approach, two statistical models, namely a 'language model' and a 'translation model' are built. This technique has been successfully used in speech recognition. A language model provides probabilities of occurrence of the sentence in the language, P(S) and a translation model provides probability of a target sentence given source sentence, P(T/S). An N-gram model is used to build the language model. Language models for both source and target languages are built. The translation model is computed using a word-level aligned bilingual corpus. For details of the modeling process, refer to Brown et al. (1990). Using language model probabilities and conditional probabilities of the translation model, P(S/T) is computed using the following formula:

$$P(S/T) = \frac{P(S)P(T/S)}{P(T)}$$

This approach does not require explicit encoding of linguistic information. On the other hand, it is heavily dependent on the availability of good quality bilingual data in very large proportions, which is currently not available for most languages (Arnold et al. 1993).

In this thesis the MT system used is based on the transfer architecture. The Lexical Functional Grammar (LFG) formalism will be used for the analyses of both languages.

2.2 Lexical Functional Grammar (LFG)

This section presents a brief overview of a linguistic formalism, LFG, which is well established for the analysis and generation modules of machine translation systems.

Lexical Functional Grammar (LFG) is a unification-based linguistic formalism which is suitable for computation purposes. LFG uses different structures for representing the following different levels of linguistic information that is contained in a sentence:

1) Constituent Structure (c-structure): a structure for representing sentence structure (Kaplan, 1989).
2) Functional Structure (f-structure): a relatively order-free attribute-value bundle pair for representing higher level syntactic and functional information (Kaplan, 1989).
3) Semantic Structure (s-structure): an f-structure look-alike structure for representing semantic information (Halvorsen and Kaplan, 1988).

Structural correspondences are defined to relate the elements of a c-structure to those of an f-structure and the elements of an f-structure to those of an s-structure (Kaplan, 1989). The following section gives an explanation of c-structures and f-structures. For a detailed explanation of s-structure see Halvorsen and Kaplan (1988).

2.2.1 Constituent Structure (c-structure)

The c-structure in LFG represents the external structure of a sentence in the form of a phrase structure tree. It shows the syntactic categories and the linear order of the constituents of the sentence. It also shows the hierarchical grouping of words in a sentence, i.e., how each phrase within the sentence is formed by the combination of words in the sentence and how these phrases combine to form the sentence itself. The hierarchical grouping of words in a sentence is governed by phrase structure rules which are represented by a context-free grammar.

A context free grammar is a mathematical system for modeling constituent structures. A context free grammar has four parameters. (Jurafsky and Martin, 2000)

1. A set of non-terminal symbols, N
2. A set of terminal symbols, ∑ which is disjoint from N
3. A set of productions P, of the form A→α where A ε N and α is a string of symbols from where α ε (∑ U N)*
4. A start symbol S where S ε N

For instance, consider the sentence "John gave him a book". The set of phrase structure rules that describes the structure of this sentence and other sentences of this form is:

(2.1)

S → NP VP
VP → V NP NP
NP → PRON | (DET) N

where S stands for 'Sentence', NP stands for 'Noun Phrase', VP stands for 'Verb Phrase', N stands for 'Noun', V stands for 'Verb' and 'DET' stands for 'Determiner'.

The c-structure of the sentence in the sentence "John gave him a book." can be obtained by applying these phrase structure rules as shown in Figure 2.3.

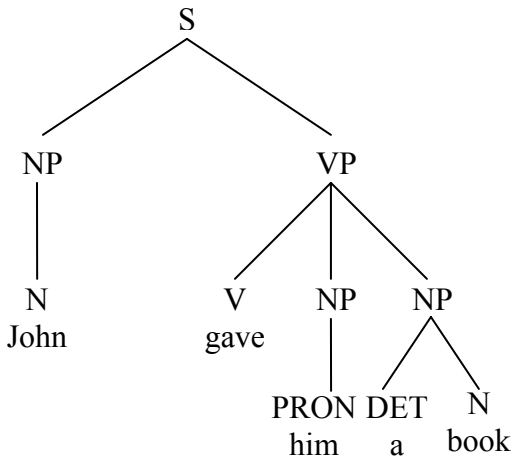

Figure 2.3: c-structure of sentence "John gave him a book"

The same tree can be shown in bracketed form as follows.

(2.2)
[S [NP[N[John]] [VP[V[gave] [NP[PRON[him]]] [NP[DET[a] [N[book]]]]

The c-structure displays information about the part-of-speech of each constituent in a sentence and the syntactic structure of the sentence. As the c-structure encodes surface syntactic information like word order and phrasal structure; it is language dependent. Although the c-structure contains information explaining how each constituent is grouped to form a sentence which can aid analysis of source language sentences or generation of target language sentences in an MT system, its language dependence only allows it to capture the shallow syntactic information of sentences. This makes it insufficient for performing the transfer of sentences from one language to another.

2.2.2 Functional Structure (f-structure)

While the c-structure captures the external structure of a sentence, the f-structure represents the internal structure of a sentence. This includes the representation of the higher syntactic and functional information of a sentence. The higher syntactic information of a sentence refers to the grammatical information of a lexical item, e.g., the word ‘cats’ is in *plural* form and the word ‘ate’ is expressed in the *past* tense. The functional information of a sentence includes information about functional relations between parts of sentences and how parts of the sentence affect each other. The f-structure also expresses information about the kind(s) of syntactic functions that each predicator (e.g. verb or preposition) governs. The higher syntactic and functional information of a sentence is represented in the f-structure as a set of attribute-value pairs. These pairs form the nodes of an acyclic graph structure. In an attribute-value pair of an f-structure, the attribute corresponds to the name of a grammatical symbol (e.g. NUM, TENSE) or a syntactic function (e.g. SUBJ, OBJ) and the value is the corresponding feature possessed by the concerned constituent. The value for each attribute can be an atomic symbol, a semantic form or a subsidiary f-structure (Kaplan, 1989). An atomic value is used to describe a grammatical feature of a constituent, e.g. the tense of a verb,

whether a noun is of a singular or plural form, etc. (2.3) is an example of an attribute-value pair with an atomic value showing the tense of the verb ‘gave’:

(2.3)

$$\left[\text{TENSE} \quad \text{PAST}\right]$$

In LFG terminology, a semantic form expresses the semantic interpretation of a predicate. This semantic interpretation is represented in terms of the syntactic functions a predicator governs. The feature representing this semantic form is termed PRED. For instance, the attribute-value pair which encodes the semantic form of the verb ‘gave’, as in “John gave him a book.” is:

(2.4)

$$\left[\text{PRED} \quad \text{'GIVE} < \left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{OBJ}\right)\left(\uparrow \text{OBJ2}\right)\right]$$

This states that the verb ‘give’ requires a subject, an object and a secondary object (OBJ2) as its arguments. A sentence using the verb ‘give’ will be considered incomplete without any one of these arguments and will be incoherent if any additional argument is present.

The functional structure of a syntactic function is encoded as a subsidiary f-structure in an attribute-value pair. For instance, the f-structure representation of the NP ‘John’ which functions as the subject in a sentence is:

(2.5)

$$\begin{bmatrix} & \begin{bmatrix} \text{PRED} & \text{'JOHN'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{SUBJ} & \end{bmatrix}$$

As an f-structure may contain subsidiary f-structure(s), so essentially the f-structure is a multi-leveled tree-like structure. Nevertheless, an f-structure is not a tree because some of the attributes that appear in different places within it can sometimes be linked with each other. Within the same level of an f-structure, the attribute-value pairs can appear in any order. Each attribute has a unique value.

As mentioned in the previous section, the c-structure of a sentence is assigned using phrase structure rules. The phrase structure rules in (2.1) shown in Section 2.2.1 did not carry any functional information about the constituents within a sentence. Thus, they are insufficient for assigning f-structures. In order to enrich the syntactic information carried by phrase structure rules, they are equipped with functional annotations. For example, (2.6) a. is only phrase structure rule for sentences whereas (2.6) b. is annotated with functional description.

(2.6)

a. S → NP VP.

b. S → NP: (↑ SUBJ) = ↓; VP: ↑ = ↓.

where the functional annotation for the NP node expresses the grammatical relation "*the f-structure which fills the value of the attribute 'subject' (SUBJ) of the mother of this NP node's is the f-structure of this NP node*"; and the functional annotation '↑ = ↓' for the VP node indicates that the functional information encoded in this VP node is passed to the f-structure of its mother node.

In addition to appearing in the form of functional annotations on phrase structure rules, most of the functional information appears in the lexical items, e.g.:

(2.7)

John: N, (↑ PRED) = 'john';
(↑ NUM) = SG;
(↑ PERS) = 3.
gave: V, (↑ PRED) = 'give <(↑ SUBJ) (↑ OBJ) (↑ OBJ2)>';
(↑ TENSE) = PAST.

The lexical items form the terminals of the grammar rules, e.g.:

(2.8)

N → John: (↑ PRED) = 'john';
(↑ NUM) = SG;
(↑ PERS) = 3.
V → gave: (↑ PRED) = 'give < (↑ SUBJ) (↑ OBJ) (↑ OBJ2)>';
(↑ TENSE) = PAST.

d whereas ↓ denotes "the subject of the s the formation an f-structure from the given annotated grammar. Each c-structure node is assigned a variable

corresponding to its f-structure. Then the '↑' and the '↓' in annotations are replaced with appropriate variable names. This process is called instantiation. Then all the equations are solved by applying unification. Unification can be described as follows (Dalrymple 2001):

- An empty feature structure is the identity element.
- The unification of an attribute with another attribute is successful if both attributes have the same value, otherwise unification will fail.
- The feature structure f1 unified with feature structure f2 makes feature structure f3 in the following manner:
    - The set of features in f3 is the union of the features of f1 and f2.
    - The value of each feature in f3 is the value of that feature in f1 unified with the value of that feature in f2.
    - Recursively traverse through the embedded feature structures if any.
    - If any unification fails, then the whole process fails.

The f-structure for a sentence is the minimal f-structure that satisfies all of the equations.

For instance, the f-structure corresponding to the sentence "John gave him a book." is shown in Figure 2.4. The grammar rules used are as follows.

(2.9)

S → NP: (↑ SUBJ) = ↓; VP: ↑ = ↓.
VP → V: ↑ = ↓; NP: (↑ OBJ2) = ↓; NP: (↑ OBJ) = ↓.
NP → PRON: ↑ = ↓; | (DET: ↑SPEC = ↓ ;) N: ↑ = ↓.

$$\begin{bmatrix} \text{PRED} & \text{'give} < (\uparrow \text{SUBJ})(\uparrow \text{OBJ})(\uparrow \text{OBJ2}) \text{'} \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'john'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'book'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \\ \text{SPEC} & [\textit{DEF} \quad \textit{NEG}] \end{bmatrix} \\ \text{OBJ2} & \begin{bmatrix} \text{PRED} & \text{'Pro'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \end{bmatrix}$$

Figure 2.4: f-structure of sentence "John gave him a book"

Verb analysis being followed in determining transfer rules in this work is discussed in the following section.

2.3 Verb Analysis Using LFG

In this section some concepts of grammatical analysis for verbs will be described.

A sentence states a relationship between some individuals, i.e. person, place, thing etc. or asserts a property of any individual. The element in the sentences which contains information about the relationship is called the predicate. The individuals or participants in the relationship are called arguments. A grammatical unit containing one predicate and its participants is called a simple sentence or a clause (Kroeger, 2005). Generally verbs carry the idea of being or action in the sentence. They provide the essential backbone of a clause since they define what arguments a clause needs to be grammatical. Verbs contribute their semantic form or predicate, i.e. PRED and subcatagorization frames. Subcatagorization can be defined as semantic roles or grammatical relations. In the analysis given below grammatical relations are used. It defines what grammatical relations a particular verb can take as argument. Moreover, verbs define the tense and aspect features of sentence.

2.3.1 Grammatical Relations

Grammatical relations are a grammatical notion and they can be defined using a grammatical criterion. In this section we will discuss grammatical properties that can be used to identify different grammatical relations. The analysis under consideration assumes the following grammatical relations: Subject (SUBJ), Object (OBJ), Secondary Object (OBJ2), COMP, and XCOMP.

2.3.1.1 Subject (SUBJ)

A noun phrase in the clause acts as the subject of the clause. In most languages the assumed rule is that all verbs subcategorize for the subject. Some properties of subjects in English are stated in Kroeger (2005) taken from Bickford (1998:43). Word order is an important clue for identifying the subject in English. The noun phrase coming before the verb is normally the subject. Verb agreement is also an indication. The verb agrees with the subject in English, e.g. a third person singular subject adds an –s to the verb. Another clue is nominative case marking on NP. This is visible only by morphology in pronouns in English. There is another test for subjects using question words. If a question word replaces the subject of the sentence, the rest of the sentence remains unchanged. But if a question word replaces some other element of the sentence, the auxiliary will move before the subject. If the sentence does not contain an auxiliary, an additional 'do' or 'did' will be added immediately after question word (Kroeger 2005).

The properties used for identifying subjects may vary from language to language. For example, different case markings can help in identification of the subject in different languages. In German the nominative case marking helps in identifying the subject. In Ergative languages, the ergative case marking also identifies the subject. An analysis for Hindi/Urdu subject identification is presented in Mohanan (1994).

2.3.1.2 Object (OBJ)

The second argument of transitive verbs is usually an object. In English, the object can be clearly identified by position. The object must follow the verb and be adjacent to it. In free order languages like German and Hindi, case marking is the indicator for object. A

noun phrase with accusative case is analyzed as an object in German. For example, the following two sentences of German have the same meaning and the accusative case helps in identifying the object of sentence (Butt et al., 1999).

(2.10) Der Fahrer startet den Traktor.
The.Nom driver starts the.Acc tractor

(2.11) Den Traktor startet der Fahrer.
the.Acc tractor starts the.Nom driver
(The driver is starting the tractor.)

Urdu is also a free order language and case markings help in identifying the object.

A cross linguistic test for objects is passivisation. When a sentence is passivised, its object becomes the subject. By this transformation we can differentiate the object from other roles mentioned in this section. For example, "the house" is the object of the sentence in Example (2.13) (Butt et al., 1999).

(2.12) He built the house (Active voice)
SUBJ OBJ

(2.13) The house was built (Passive voice)
SUBJ

If the noun phrase is not the object, it cannot be passivised. For example, as 'home' is not the object in the sentence mentioned in (2.15), the passive construction in (2.16) is not valid.

(2.14) He went home. (Active Voice)
SUBJ OBL

(2.15) *home was gone.

2.3.1.3 Oblique (OBL)

Obliques are arguments which are not of an appropriate morphosyntatic form to be considered objects and which do not undergo the syntactic processes which affect objects, such as passivisation (Butt et al., 1999). These arguments are associated with a particular semantic role and are marked to indicate their role. For example 'to him' in the sentence 'I gave a book to him' is OBL_{GOAL} (Dalrymple, 2001).

Obliques are generally prepositional phrases (PP), or in some languages case marked noun phrases as discussed by Nordlinger (1998) (Dalrymple, 2001). Noun phrases can also be obliques, sometimes with a changed morphological form. For example, 'home' in sentence (2.15) is an oblique. We can see it cannot be passivised.

2.3.1.4 Secondary Object (OBJtheta)

In many languages, a clause may contain more than one noun phrases as the object. In English, ditransitive verbs such as 'give' subcategorize for a subject, an object and a secondary object. The primary object can be identified by its position in English. It must be adjacent to the verb followed by the secondary object. The secondary object is known as OBJtheta (Dalrymple, 2001).

(2.16) She gave him a book

SUBJ OBJ OBJtheta

In sentence (2.16) 'him' is the OBJ and 'book' is the OBJtheta.

In German the dative case of an NP can be a clue for identifying the OBJtheta but there are some constructions where the dative case occurs but the NP is not an OBJtheta. Thus the OBJtheta can be identified by adding one more condition that there should be a

primary object also present in the clause. (Butt et al., 1999). Similarly for Urdu the dative case distinguishes OBJtheta from OBJ as shown below.

(2.17) اس نے اسے کتاب دی۔

[us ne]	usey	kitab	di.
Erg.3.sg	Dat.3.sg	Nom.3.sg.F	3.sg.F
SUBJ	OBJtheta	OBJ	

2.3.1.5 XCOMP and COMP

A clause can also be the argument of a verb. For example in the following sentences the verbs 'want' and 'know' take clauses as arguments.

(2.18) I want [to do this work].

(2.19) I know [that this tractor is red].

Such arguments can be categorized as an XCOMP and a COMP. An XCOMP is a complement whose subject lies outside the clause, as in (2.18), "I" is logically the subject of the clause "to do this work" but it is not within the clause. Such a clause is known as an infinite clause. A COMP is a closed complement with its own subject as in (2.19) "the tractor", the subject of the clause, is within the clause. This is known as a finite clause.

2.3.1.6 Adjunct

Prepositional phrases and adverbs which are not included in the subcategorization frame of the verb are considered adjuncts. For example, in sentence (2.20), 'last night' is an adjunct.

(2.20) She met him last night

SUBJ OBJ ADJUNCT

These phrases are optional in the sentence and are added to convey additional information such as the time and place of event. One indication of a phrase being an ADJUNCT is that it is always optional. Secondly, an unlimited number of ADJUNCTs

can be added to sentence whereas an argument can occur only once. Moreover, they can be freely added to most clauses, whereas the arguments of a verb are restricted by the type of verb (Kroeger, 2005).

2.4 Transfer Problems

In this section, problems in lexical transfer will be discussed. Then a classification of lexical-semantic divergences presented by Dorr (1994) will be presented and their Hindi examples will be shown which are taken from Gupta and Chatterjee (2001).

Lexical transfer refers to changing source language words into target language words. In transfer at f-structure level, this issue is realized as choosing the correct value of the PRED feature and applying appropriate structural changes if required. For choosing the correct translation, a clue from the sentence structure or surrounding context needs to be examined. For example the English word ‘know’ can be translated as ‘connaitre’ or ‘savoir’ in French. The choice of these two translations depends roughly on whether the word subcategorizes for the noun phrase as an object or a complement. The following sentences illustrate the point. (Hutchins and Somers 1992, p. 100)

(2.21) I know the right answer.

Je connais la bonne résponse.

(2.22) I know what the right answer is.

Je sais quelle est la bonne résponse.

Another example can be the translation of the word ‘eat’ in German. It is translated to ‘essen’ if the subject is human, and to ‘fressen’ otherwise. Some cases are more difficult, for example the word ‘library’ is translated as ‘bibliothek’ if it is part of an academic or research institute but ‘bucherei’ if it is a public library. In this case the selection of a translation is not easy. (Hutchins and Somers 1992)

Another problem in transfer is when the structure of the translation does not match the source sentence. Dorr (1994) presented some major lexical-semantic divergence problems which are valid cross linguistically. The divergence classes and their examples presented by Dorr (1994) are as follows.

1. *Thematic divergence*

The theme of the sentence is the subject in the source language which changes into an object in the target language. For example when the following English sentence (2.23) is translated into Spanish, the theme of the sentence "Mary" changes from object to subject.

(2.23) I like Mary
Maria me gusta a mi
(Mary pleases me)

2. *Promotional divergence*

This is a head switching problem. An adverbial phrase in the source language changes into the main verb in the target language. For example, in the Spanish translation of the English sentence in (2.24), the adverb is changed into the main verb. A similar problem can be seen in the Hindi translation in (2.25).

(2.24) John usually goes home
Juan suele ir a casa
(John tends to go home)

(2.25) The fan is on
pankhaa chal rahaa hai

3. *Demotional divergence*

This is opposite to promotional divergence. The main verb in the source language is changed into an adverbial phrase in the target language. One example of it can be seen in the following English to German translation.

(2.26) I like eating

Ich esse gern

(I eat likingly)

4. Structural divergence

The noun phrase is converted to a prepositional phrase. For example, in the English to Spanish translation of (2.27) "the house" is converted to "en la casa", a PP. A similar problem can be seen in the Hindi translation in (2.28).

(2.27) John entered the house

"Juan entro en la casa"

(John entered in the house)

(2.28) Ram attended the meeting

ram sabha mai upashtit tha

5. Conflatational divergence

Sometimes in the source language one word is used to explain a concept and its argument such that it is contained within the word, whereas the target language requires an explicit argument. For example the English word 'stabbed' is usually not translated in one word in other languages. This can be seen in the following English to Spanish and English to Hindi Examples.

(2.29) I stabbed John

Yo le di punaladas a Juan

(I gave knife-wounds to John)

(2.30) He stabbed me

usne mujhe chaaku se maaraa

(He hit me with knife)

6. Categorical divergence

In this case the category of predicates is changed. For example, the adjective 'hungry' in the English sentence (2.31) changes into the noun in the German translation. A similar problem can be seen in the following Hindi translation.

(2.31) I am hungry

Ich habe Hunger

(I have hunger)

(2.32) I am feeling hungry

mujhe bhukh lag rahii hai

7. Lexical divergence

Sometimes the word being used in the target language is different in literal meaning as compared to the source language, e.g., when the English sentence (2.33) is translated into Spanish, the verb 'forzo' (force) is used instead of 'break'. A similar problem can be seen in the following Hindi translation.

(2.33) John broke into the room

Juan forzo la entrada al cuarto"

(John forced (the) entry to the room)

(2.34) They run into the room

woye daurte huye kamre mein ghus gaye

3 Problem Statement

The problem statement of this thesis is:

"*To identify the issues in the lexical transfer of verbs for an English to Urdu machine translation system and to present their solutions in the form of lexical transfer rules*."

The different aspects of this statement are explained in the following sections.

3.1 Need

As has been shown in section 2.4, the transfer process in a machine translation system is not a one to one word replacement problem. English to Urdu transfer based systems need transfer rules to pick the correct translation for words and to make structural changes if required. Since verbs are the backbone of any sentence, transfer rules relevant to verbs are of great importance.

3.2 Scope

Transfer rules can be written at various levels of analysis. One possibility is to use semantic level representation. Such representation requires deep analysis of source and target languages. Another possibility is to use f-structure representation. This requires less analysis time and still provides a language independent representation. Verb transfer rules in this thesis will be identified at the f-structure level. 900 verbs will be analyzed and their lexicon will be developed. By the analysis of these verbs, templates of changes which occur during translation will be identified

In addition to meaning, verbs also indicate the tense aspect of a sentence. It is mostly a structural transfer issue, i.e., it does not depend on the individual verb. Tense aspect issues will not be addressed unless there is some influence on the verb analysis at the lexical level.

4 Methodology

Firstly, grammatical analysis of English and Urdu was required so that on the basis of that structural mismatches could be identified. The Machine Translation project, from now onwards referred to as the MT Project, is being conducted under the Urdu Localization Project of EGD (E-Government Directorate) at CRULP. The work done on grammar analysis under the MT Project is used as the basis of the issues presented in the thesis.

Next, a variety of verbs and their translations were selected for analysis. The verbs and their translations for this study were taken from work done under the MT Project. These verbs were selected on the basis of frequency, taken from the British National Corpus (BNC) and were translated using different English to Urdu dictionaries and native speaker knowledge. Different senses and subcategorization frames were analyzed and their translations were done.

These verbs were analyzed to find divergences in English and their Urdu translations. Approximately 900 verbs were analyzed. Categories were defined on the basis of different transformations required for translation. Generalized rules for these categories were defined. Major categories are discussed in the thesis. The rules are realized using the MT system mapper and problems faced during implementation of the rules are discussed.

The transfer system developed under the MT project is used to realize the transformation rules found during analysis. A description of the working of the transfer system is as follows:

(The rules in this section may be dummy rules, created specifically for the purpose of illustration.)

The transfer system uses mapping rules to define transfer behavior for features i.e. attribute value pairs of f-structures. All the rules are uni-directional. For example, to create a new structure, SUBJ, in Urdu corresponding to a SUBJ structure in English, the following rule is used.

(4.1)

```
SUBJ
[
        --> SUBJ;
]
```

Any structure can be removed from the target structure along with its child features by using a NULL operator. For example the article 'the' of English does not get translated into Urdu and is eliminated it using rule 4.2. Rules may also have a condition to choose the context in which the rule should be applied.

(4.2)

```
DET
[
        (DEF =c {POS}) --> NULL:
]
```

A feature can be assigned the same value that it had in the source structure, or it can be given a new value, using Rule 4.3 and 4.4 respectively.

(4.3)

```
NUM
[
        --> NUM = NUM:
]
```

(4.4)

```
NUM
[
        --> NUM = SG:
]
```

There may be multiple rules for each feature. This is shown in the following rule block.

(4.5)

```
CONJ_FORM
[
        (CONJ_FORM =c {OR}) --> CONJ_FORM = {YA};
        (CONJ_FORM =c {AND}) --> CONJ_FORM = {AUR};
]
```

The rule is assigning the value YA to the CONJ_FORM feature in Urdu if there is a CONJ_FORM with the value OR in English. In the same way, the next rule is assigning an AUR for each AND in English.

If there is no rule for any feature, that feature is skipped in the target structure.

The rules are applied in the order that features are present in the f-structure. The f-structure is traversed and an appropriate rule is found and fired for each feature.

A bilingual mapping lexicon is used to map predicates in the structure. The syntax of the lexicon is the same as that of the rules for features. The only difference is that each lexicon entry corresponds to an English predicate whereas in the feature rules each entry corresponds to a feature.

In a one to one mapping of words, the rule simply states the corresponding Urdu predicate. For example, the following rule gives the translation of the noun ‘book’.

(4.6)

```
book_n
[
        --> PRED = ‘kitab’;
]
```

To handle complex translations, multiple actions can be defined on the right side of the rule. Any feature or structure can be added as required. For example, in the following rule, the noun ‘chemical’ has a two word translation made up of the noun ‘mada’ and adjective modifier ‘kimyai’.

(4.7)

```
chemical_n
[
        --> PRED = ‘mada’, ADJUNCT.ADJ = ‘kimyai’;
]
```

The phenomenon occurring in the above mentioned rule can be captured and represented in a generalized form so that it can be re-used in any translation where it occurs again. This is done as shown in the following rule. This generalized rule is called macro.

(4.8)

```
#NOUN_WITH_ADJ(arg1, arg2): PRED = arg1, ADJUNCT.ADJ =
```

Using this, the rule for the word ‘chemical’ can be represented as follows.

(4.9)

```
chemical_n
[
        --> NOUN_WITH_ADJ(‘mada’, ‘kimyai’);
]
```

This rule will be preprocessed to produce the original rule (4.7) in its expanded form before it is eventually used by the MT system for mapping purposes. Expanded rules of this form will be referred to as instantiated rules from now onwards.

If any rule for any structure or feature needs to be changed in the context of a particular word, the rule is overridden in the lexical rule of that word. For example, the noun ‘grammar’ is translated to ‘kewaid’ in Urdu, which is the plural of ‘kaida’. Following is the rule for word ‘grammar’.

(4.10)

```
grammar_n
[
        --> PRED = 'kaida', INSERT(NUM, NUM_PL),
NUM_CHANGE = {TRUE};

]
```

The NUM feature for ‘grammar’ is singular but it needs to be plural for Urdu. So the default rule for the NUM feature can be overridden with a new rule, NUM_PL. This new rule is defined as follows.

(4.11)

```
NUM_PL
[
        (t::NUM_CHANGE =c {TRUE}) --> NUM = {PL};
]
```

The above mentioned rule (4.11) instructs the mapper to make a NUM feature with the value PL in the target structure when there is a NUM_CHANGE with the value TRUE in the target structure. The t:: indicates that the target structure should be searched for the feature NUM_CHANGE. This new rule will be added to the rule block for NUM and will be used when the condition to apply this rule is true, i.e., the NUM feature of the word ‘grammar’ is being transferred. To add this rule, an operator, INSERT, is used, which takes the name of the rule block to which the rule will be inserted and the name of the new rule as operands. The scope of this modification is within the sentence.

5 Results

In this section the transformations are discussed which are identified in the English to Urdu translation of verbs. A major factor causing these transformations is the phenomena of complex predication (Mohanan, 1994) in Urdu. English verbs are very frequently translated into a noun and verb combination in Urdu causing transformation in arguments of the verb at the Urdu end. The first two sections discuss these transformations. XCOMP conversion section describes noun like analysis of infinitive phrase in Urdu and rules required to map English XCOMP to Urdu analysis. Some already known phenomena such as structural divergence and conflatational divergence (Dorr 1994) are also discussed in OBJ/OBL Conversion and OBL/ADJUNCT Insertion. Ditransitive Conversion rules are described to handle mismatch of ditransitive verb analysis in English and Urdu.

5.1 Verbal Noun Conversion

Many times English verbs get translated to a noun and verb combination in Urdu. These noun-verb combinations are analyzed as complex predicates (Mohanan, 1994) in Urdu. The noun in this case is not considered as argument of the clause, rather the noun and verb combine to form the predicate of the clause. To illustrate this we take the following example of the verb ‘invent’.

(5.1) a.

Nadia	invented	the	design.
N	V		N
Nom.3.sg.F			Acc.3.sg.M
SUBJ		OBJ	

b. نادیہ نے ڈیزائن ایجاد کیا۔

[Nadia ne]	design	ijaad	kia.	
N CM	N	N	V	
Erg.3.sg.F	Nom.3.sg.M	Nom.3.sg.F	3.sg.M	
SUBJ	OBJ			(Active

Voice)

c. نادیہ نے ڈیزائن کو ایجاد کیا۔

[Nadia ne]	[design ko]	ijaad	kia.
N CM	N CM	N	V
Erg.3.sg.F	Acc.3.sg.M	Nom.3.sg.F	3.sg.M
SUBJ	OBJ		(Active Voice)

d. ڈیزائن ایجاد کیا گیا۔

design	ijaad	kia	geya
N	N	V	PASS_AUX
Nom.3.sg.M	Nom.3.sg.F	3.sg.M	3.sg.M
SUBJ			(Passive Voice)

In the example (5.1) b, we can see that verb is showing masculine agreement which is with 'design' not with 'ijad' which is feminine. Moreover, in (d) 'design' became the subject of the sentence when passivised. These facts show that the noun added as translation of the verb is not acting as an object. For further detailed discussion on this see Mohanan (1994).

To model this phenomenon in the MT system, the noun is added as the head of a clause along with a subcategorization frame. The verb coming with it is indicated as feature 'ACTION_TYPE'.

The rule is realized as follows.

Rule - 1.

#VERBAL_NOUN(agr1, arg2): PRED = arg1, PRED.GF = PRED.GF, ACTION_TYPE = agr2;

Rule - 1 is used in the verb 'invent' as shown below.

```
invent_v
[
        --> #VERBAL_NOUN(‘ijad’, kar) ;
]

invent_v
[
        --> PRED = ‘ijaad’, PRED.GF = PRED.GF, ACTION_TYPE = kar;
]
```

387 English verbs were found with verbal noun translations. These verbs are listed in Appendix A.1 along with their translations.

5.2 Object Insertion

This section discusses various cases of the phenomenon of object insertion. This transformation category is very similar to verbal noun case mentioned in section 5.1. It also results in a noun-verb combination, but the noun in this case is considered as the object of the clause. This phenomenon affects the subcategorization frame of the translated sentence.

5.2.1 Intransitive Verb

The simplest case of object insertion can be observed in intransitive verbs. We start the discussion by illustrating an example of intransitive verb of English translated into noun and verb in Urdu.

(5.2) a. He whispered.

PRON V
Nom.3.sg.M
SUBJ

b. اس نے سرگوشی کی۔

[Us ne] sargoshi ki

PRON CM N: Whisper V: Do
Erg.3.sg.M Nom. 3.sg.F 3.sg.F
SUBJ OBJ Active Voice

c. سرگوشی کی گئی۔

sargoshi	ki	gayi	
N: Whisper Nom. 3.sg.F SUBJ	V: Do 3.sg.F	PASS_AUX 3.sg.F	Passive Voice

In Sentence (5.2), the English verb 'whisper' is translated into a combination of Urdu noun 'sargoshi' and Urdu verb 'ki'. The grammatical function of the noun 'sargoshi' is determined as object. This analysis is made clear by the fact that the verb 'ki' shows agreement with 'sargoshi', which is feminine. This agreement is regular according to the rule of Urdu that when a subject is case marked, the verb agrees with the object (Mohanan, 1994). Sentence (5.2) c also confirms this analysis by showing that 'sargoshi' is the subject of the passive sentence.

While translating such verbs, an additional argument "object" will be introduced and the Urdu translation will have a transitive frame.

The figure 5.1 illustrates change in subcategorization from English to Urdu.

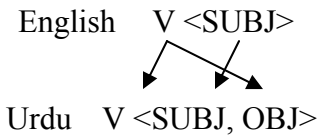

Figure 5.1: Subcategorization transfer for Rule - 2

In the MT Project, this rule is dealt with as follows.

Rule - 2. (preliminary version)

#TRANS(arg1,arg2) : PRED = arg1, PRED.GF = <SUBJ,OBJ>, OBJ.PRED = arg2;

arg1 is the translated verb in Urdu. The English subcategorization frame is overridden by the new frame, <SUBJ,OBJ> and arg2 is added as an object.

This rule will be used in the verb ‘whisper’ as follows.

whisper_v
[
(PRED.GF =c {<SUBJ>}) --> #TRANS (‘ker’, ‘sergoshi’);
]
Instantiated Rule:
whisper_v
[
(PRED.GF =c {<SUBJ>}) --> PRED = ‘ker’, PRED.GF = <SUBJ,OBJ>, OBJ.PRED = *‘sergoshi’*;
]

This rule is generically applicable to other intransitive frames. Following are examples of some other intransitive verbs having different subcategorization frames, exhibiting the same phenomenon.

(5.3) a.

I	communicated	[with	Nadia].
PRON	V	P	N
Nom.1.sg			Acc.3.sg.F
SUBJ		OBL	

b. میں نے نادیہ سے بات کی۔

[men ne]	[nadia se]		bat	ki.
PRON CM	N	P	N	V: Do
Erg.1.sg.M	3.sg.F		Nom.3.sg.F	3.sg.F
SUBJ	OBL		OBJ	

In the above sentence (5.3) b, the oblique remains unaffected; it is a prepositional phrase in both sentences. A noun is added as the object in the Urdu translation. Another example sentence containing SUBJ and COMP in English is shown in (5.4).

(5.4) a.

He answered	[that	he	will	come].
PRON V	CONJ	PRON	AUX	V
Nom.3.sg.M		Nom.3.sg.M		
SUBJ	COMP			

b. اس نے جواب دیا کہ وہ آئے گا۔

[Us ne]	jewab	diya	[ke wo aye ga]
PRON CM	N	V	CONJ PRON V AUX
Erg.3.sg.M	Nom.3.sg.M	Nom.3.sg.M	
SUBJ	OBJ		COMP

Keeping in view the above mentioned examples, the rule can be generalized as follows.

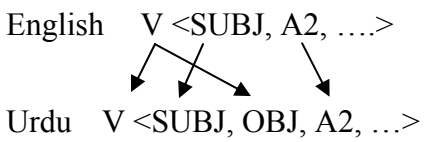

Figure 5.2: Subcategorization transfer for generalized Rule - 2

Rule - 2. (generalized version)
#TRANS(arg1,arg2,agr3) : PRED = arg1, PRED.GF = arg3, OBJ.PRED = arg2;

As more than one word is being used in the target language to translate the source language verb, this transformation can come under the category of Conflatational Transformation of Dorr (1994).

Verbs found exhibiting this phenomenon are listed in Appendix A.2 along with their translations. There were 129 such verbs.

5.2.2 Transitive verbs

We have seen object insertion in verbs in which there was no object present in English. In this section we will discuss the same phenomenon for transitive verbs.

When the verb to translate is transitive, there are two potential candidates for the object in the Urdu translation; one is the original one coming from English and the other is the one produced as a result of MT rule. The first question to answer is what the object in the

translated sentence should be. Next we will see what happens to the additional argument. In the following sections we will discuss two transformations for such verbs.

5.2.2.1 Object to Secondary Object

One transformation that is noticed in English to Urdu translation is the conversion of the object into a secondary object in translation. We take a transitive verb with a noun-verb translation as an example (5.5).

(5.5) a.

Nadia	answered	him.
N	V	PRON
Nom.3.sg.F		Acc.3.sg.M
SUBJ		OBJ

b. نادیہ نے اسے جواب دیا۔

[Nadia ne]	usey	jewab	diya.	
N CM	PRON	N	V	
Erg.3.sg.F	Dat.3.sg	Nom.3.sg.M	3.sg.M	
SUBJ	OBJ2	OBJ		(Active voice)

c. اسے جواب دیا گیا۔

usey	jewab	diya	geya	
PRON	N	V:GIVE	PASS_AUX	
Dat.3.sg	Nom.3.sg.M	3.sg.M	3.sg.M	
OBJ2	SUBJ			(Passive voice)

In sentence (5.5) b, the verb ‘answer’ is translated to ‘jewab dena’. ‘jewab’ is the noun added as part of translation of the verb. It is analyzed as object because of its agreement with the verb and because it is acting as the subject in the passivised form of the sentence (5.5) c.

Now the object of the English sentence needs to be transformed into some other role to accommodate the newly inserted object of Urdu translation. As we see in the above

mentioned example (5.5) b, the English object is analyzed as a secondary object in the Urdu translation. It is marked with a dative case marker which is regular for secondary objects and the verb ‘dena’ is a regular ditransitive verb.

Figure 5.3 illustrates the change in subcategorization from English to Urdu.

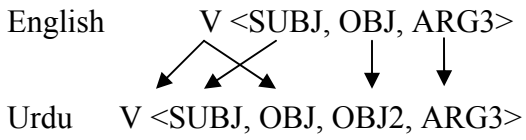

Figure 5.3: Subcategorization transfer for Rule - 3

This rule is realized in the MT system as follows:

Rule - 3.

#DITRANS(arg1, arg2,arg3): PRED = arg1, PRED.GF = arg3, OBJ.PRED = arg2, INSERT(OBJ, OBJ_OBJ2), DITRANS_FLAG = {TRUE};

Corresponding Insert Rule

```
OBJ_OBJ2
[
        (DITRANS_FLAG =c {TRUE}) --> OBJ2;
]
```

In this transformation, the rule adds an object and replaces the default rule for object to create OBJ2 instead of OBJ.

The use of this rule for the verb ‘answer’ will be as follows:

answer_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> #DITRANS (‘dey’, ‘jewab’, <SUBJ,OBJ,OBJ2>);
]
Instantiated Rule:
answer
[
(PRED.GF =c {<SUBJ, OBJ>}) --> PRED = ‘dey’, PRED.GF = <SUBJ, OBJ, OBJ2>, OBJ.PRED = ‘jewab’, INSERT(OBJ, OBJ_OBJ2), DITRANS_FLAG = {TRUE};
]

The English verbs in which this kind of conversion is seen are listed in Appendix A.3 along with their translations.

5.2.2.2 Object to Genitive Modifier

In the above section we have seen examples of transitive verbs with noun-verb translation and their analysis. Following is another example where the verb ‘help’ gets translated into the noun ‘meded’ and the verb ‘kerna’.

(5.6) a.

Nadia	helped	him.
N Nom.3.sg.F SUBJ	V	PRON Acc.3.sg.M OBJ

b. نادیہ نے اس کی مدد کی۔

[Nadia ne]	[use ki	meded]	ki.	
N CM Erg.3.sg.F SUBJ	PRON CM Gen.3.sg.F	N Nom.3.sg.F OBJ	V 3.sg.F	(Active voice)

c. اس کی مدد کی گئی۔

[use ki	meded]	ki	gayi

PRON	N	V	PASS_AUX
Gen.3.sg.F	Nom.3.sg.F	3.sg.F	3.sg.F
SUBJ		OBJ	(Passive voice)

In the example (5.6), 'meded' can be seen as the object of the Urdu sentence by the agreement and passivisation test. The object of the English sentence is translated with a genitive marker in Urdu. This genitive phrase is analyzed as a modifier of the newly added object.

Figure 5.4 explains the rule.

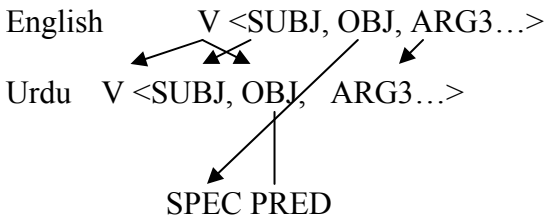

Figure 5.4: Subcategorization transfer for Rule - 4

In this transformation, the rule adds an object and replaces default rule for OBJ to make the object of the English sentence a modifier of the Urdu object.

Rule - 4.

```
#OBJ_WITH_GEN(arg1, arg2,agr3): PRED = arg1, PRED.GF = arg3,
OBJ.PRED = arg2, INSERT(OBJ,OBJ_OBJ_GEN),
OBJ_WITH_GEN_FLAG = {TRUE};
```

Corresponding Insert Rule

```
OBJ_OBJ_GEN
[
        (t:: OBJ_WITH_GEN_FLAG) =c {TRUE}) --> OBJ.SPEC.GEN;
]
```

In MT system this rule is realized for the verb 'help' as follows:

help_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> #OBJ_WITH_GEN(‘dey’, ‘jewab’,<SUBJ,OBJ>);
]
Instantiated Rule:
help_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> PRED = ‘kerna’, PRED.GF = <SUBJ,OBJ>,OBJ.PRED = ‘meded’, INSERT(OBJ,OBJ_OBJ_GEN);
]

Verbs found exhibiting this phenomenon are listed in Appendix A.4 along with their translations. 198 such verbs were found.

5.2.2.3 Object to OBL

Another variation of object insertion is English object becoming a prepositional phrase. In the following example where the verb ‘consult’ is translated into the noun ‘meshwera’ and the verb ‘kerna’.

(5.7) a.

Nadia	consulted	him.
N Nom.3.sg.F SUBJ	V	PRON Acc.3.sg.M OBJ

b. نادیہ نے اس سے مشورہ کیا۔

[Nadia ne]	[use se]	meshwera	kia.
N CM Erg.3.sg.F SUBJ	PRON P 3.sg.F OBL	N Nom.3.sg.M OBJ	V 3.sg.M (Active voice)

c. اس سے مشورہ کیا گیا۔

[use se]	meshwera	kia	gaya
PRON P 3.sg	N Nom.3.sg.M	V 3.sg.M	PASS_AUX 3.sg.M

OBL SUBJ (Passive voice)

In the above example, ‘meshwera’ can be seen as the object of the Urdu sentence by the agreement and passivisation test. The object of the English sentence is translated into a prepositional phrase in Urdu. This prepositional phrase is analyzed as OBL in the clause.

Figure 5.5 explains the rule.

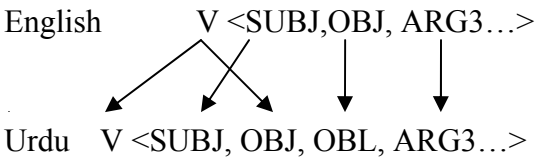

Figure 5.5: Subcategorization transfer for Rule - 5

In this transformation, the rule adds an object and replaces default rule for OBJ to make the object of the English sentence OBL of the Urdu sentence.

Rule - 5.

#OBJ_WITH_OBL(arg1, arg2,arg3,arg4): PRED = arg1, PRED.GF = arg4, OBJ.PRED = arg2, OBL.PRED=arg3, OBL.PRED.GF = <OBJ>,INSERT(OBJ,OBJ_OBJ_OBL), OBJ_WITH_OBL_FLAG = {TRUE};

Corresponding Insert Rule

```
OBJ_OBJ_OBL
[
        (t:: OBJ_WITH_OBL_FLAG) =c {TRUE}) --> OBL.OBJ;
]
```

In the MT system this rule is realized for the verb ‘consult’ as follows:

consult_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> #OBJ_WITH_OBL(‘ker’, ‘meshwera’,’se’,<SUBJ,OBJ,OBL>);
]

Instantiated Rule:

consult_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> PRED = ‘ker’, PRED.GF = {<SUBJ,OBJ,OBL>}, OBJ.PRED = ‘meshwera’, OBL.PRED=’se’, ADJUCT.PP.PRED.GF = <OBJ>,INSERT(OBJ,OBJ_OBJ_OBL), OBJ_WITH_OBL_FLAG = {TRUE};
]

Verbs found exhibiting this phenomenon are listed in Appendix A.5 along with their translations. A total of 39 such verbs was found.

5.2.2.4 Passivisation

We have discussed object insertion for transitive verbs in section 5.2.2. In all this discussion we considered active voice sentences. Now let’s look into passive variations of the same sentences.

(5.8) a.

He	was	answered.
PRON	AUX	V
Nom.3.sg.M		
SUBJ		

b. اسے جواب دیا گیا۔

usey	jewab	diya	geya
PRON	N	V:GIVE	PASS_AUX
Dat.3.sg	Nom.3.sg.M	3.sg.M	3.sg.M
OBJ2	SUBJ		

(5.9) a.

He	was	helped.
PRON	AUX	V
Nom.3.sg.M		

SUBJ

c. اس کی مدد کی گئی۔

[use ki	meded]	ki	gayi
PRON	N	V	PASS_AUX
Gen.3.sg.F	Nom.3.sg.F	3.sg.F	3.sg.F
SUBJ		OBJ	

For the above sentence, we can see that there are similar transformation as of Rule 3 and Rule 4, being applied on subjects of passive sentences. From this we can say that the Rule 3 and Rule 4 which were applied on 'surface' objects should actually be applied on 'deep' object of the clause. Since surface objects and deep objects are same for active voice sentences, the rules are valid for active voice but not for passive voice. So in the current system we have to write two separate independent rules for active and passive constructions for logically the same transformations. To make active and passive rules coherent, one solution could be analysis for deep grammatical relations. But opting for this solution will result in complex grammars for analysis and generation as agreement rules are followed on the basis of surface relations.

5.3 OBJ / OBL Conversion

5.3.1 OBJ to OBL

Sometimes the object in English is converted to a prepositional phrase in Urdu, which is analyzed as oblique instead of object. This transformation is referred as structural divergence in Dorr (1994).

(5.10) a.

He	met	Ali.
PRON	V	N
Nom.3.sg.M		Acc.3.sg.M
SUBJ		OBJ

b. وہ علی سے ملا۔

wo	ali	se	mila.
PRON	N	P	V
Nom.3.sg	Nom.3.sg.M		3.sg.M
SUBJ	OBL		

The following Figure 5.6 explains this rule.

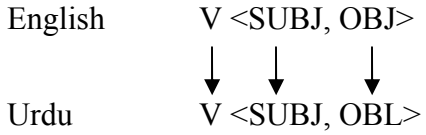

Figure 5.6: Subcategorization transfer for Rule - 6

This rule is realized in the MT system as follows:

Rule - 6.

#OBJ_TO_OBL(arg1, arg2): PRED = arg1, PRED.GF = <SUBJ,OBL>, OBL.PRED = arg2, OBL.GF = <OBJ>, INSERT(OBJ, OBJ_OBLOBJ), OBLOBJ_FLAG = {TRUE};

Corresponding Insert Rule

```
OBJ_OBLOBJ
[
        (OBLOBJ _FLAG =c {TRUE}) --> OBL.OBJ;
]
```

In this transformation, the rule adds a new preposition as head of OBL and replaces the default rule for OBJ to make it OBJ of prepositional phrase acting as OBL.

The use of this rule for the verb ‘meet’ will be as follows:

meet_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> #OBJ_TO_OBL (‘mil’, ‘se’);
]

Instantiated Rule:

meet_v
[
(PRED.GF =c {<SUBJ, OBJ>}) --> PRED = ‘mil’, PRED.GF = <SUBJ,OBL>, OBL.PRED = ‘se’, OBL.GF = <OBJ>, INSERT(OBJ, OBJ_OBLOBJ), OBLOBJ_FLAG = {TRUE};
]

A list of such verbs is as follows:

Table 5.1: OBJ to OBL Verb List

Verbs	Translation	Verbs	Translation	Verbs	Translation
pass	سے گزرنا	tackle	سے نمٹنا	meet	سے ملنا
miss	سے بچنا	hit	سے ٹکرانا	reach	تک پہنچنا
avoid	سے بچنا	enter	میں آنا	approach	تک پہنچنا
beg	سے مانگنا	climb	پر چڑھنا	question	سے پوچھنا
fetch	میں بکنا	regret	پر پچھتانا	suit	پر جچنا
prompt	سے پوچھنا	undergo	سے گزرنا	total	تک پہنچنا

5.3.2 OBL to OBJ

Sometimes the object in the English is converted to prepositional phrase in Urdu, which is analyzed as oblique instead of object. This transformation is referred as structural divergence in Dorr (1994).

(5.11) a.

He	searched	for	a	book.
PRON Nom.3.sg.M SUBJ	V	P	ART OBL	N Acc.3.sg

b. اس نے کتاب ڈھونڈی۔

us ne kitab dhondi.

PRON CM	N	V
Erg.3.sg	Nom.3.sg.F	3.sg.F
SUBJ	OBJ	

The following Figure 5.7 explains this rule.

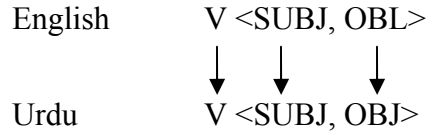

Figure 5.7: Subcategorization transfer for Rule - 7

This rule is realized in the MT system as follows.

Rule - 9.

#OBL_TO_OBJ(arg1, arg2): PRED = arg1, PRED.GF = arg2, INSERT(OBL, SKIP), INSERT(arg3,SKIP), SKIP_FLAG = {TRUE};

Corresponding Insert Rule

```
SKIP
[
        (SKIP _FLAG =c {TRUE}) --> ;
]
```

In this transformation, the rule skips OBL rule and translation of preposition of OBL which makes OBL of English sentence OBJ in Urdu.

The use of this rule for the verb 'search' will be as follows.

```
search_v
[
        (PRED.GF =c {<SUBJ, OBJ>}) --> #OBL_TO_OBJ ('dhond');
]
```

Instantiated Rule:

```
search_v
[
        (PRED.GF =c {<SUBJ, OBJ>}) --> PRED = 'dhond', PRED.GF =
        <SUBJ,OBJ>, INSERT(OBJ, OBJ_OBLOBJ), OBLOBJ_FLAG =
        {TRUE};
]
```

Following two verbs were found showing this transformation.

Table 5.2: OBL to OBJ Verb List

search	ڈھونڈنا
jump	پھلانگنا

5.4 XCOMP Conversion

This section describes Urdu grammar perspective for XCOMP and transformations needed to transfer XCOMP from English to Urdu. As we have described in section 2.3.1.5, XCOMP is an infinite clause as argument. Status of COMP and XCOMP is being questioned in LFG community. (Dalrymple and Lødrup, 2000; Alsina et al., 2005; Berman, 2006) In MT system, English grammar followed traditional analysis of COMP/XCOMP whereas for Urdu grammar it was decided that XCOMP should be eliminated from the analysis. In Urdu grammar, infinitive verb is treated as noun as it can appear at noun places, can take case marking and agree with verb in some cases (Butt, 1995). This decision affected the transfer rules for XCOMP and some rules were added to map XCOMP to its respective role in English. These rules are discussed in this section.

To show the behavior of XCOMP in translation, we start with a verb having SUBJ and XCOMP and a single word verb translation, which is the verb ‘want’. The English word ‘want’ is translated into ‘chahna’. Let us look into the sentence in (5.12).

(5.12) a.

He	wanted	to	fly.
PRON Nom.3.sg.M SUBJ	V	INF XCOMP	V

b. اس نے اڑنا چاہا۔

[us ne]	urna	chaha.
PRON CM Erg.3.sg.M SUBJ	V inf OBJ	V 3.sg.M

In the above sentence (5.12) b, ‘urna’ is analyzed as OBJ instead of XCOMP. So the default rule for XCOMP is as follows.

Rule - 8.

XCOMP
[
--> OBJ;
]

Now we take an example of verb having object insertion.

(5.13) a.

He	tried	to	fly.
PRON	V	INF	V
Nom.3.sg.M4			
SUBJ		XCOMP	

b. اس نے اڑنے کی کوشش کی۔

[us ne]	[[urney ki]		koshish]	ki.
PRON CM	V	CM	N	V
Erg.3.sg.M	inf	Gen	Nom.3.sg.F	3.sg.F
SUBJ			OBJ	

In the above example we can see that the infinitive verb is coming with the genitive marker ‘ki’. It is treated as a genitive modifier of the object ‘koshish’.

(5.14) a.

He	preferred	to	fly.
PRON	V	INF	V
Nom.3.sg.M			
SUBJ		XCOMP	

b. اس نے اڑنے کو ترجیح دی۔

[us ne]	[urney ko]		terjih	di.
PRON CM	V	CM	N	V
Erg.3.sg.M	inf	Dat	Nom.3.sg.F	3.sg.F
SUBJ	OBJ2		OBJ	

In sentence (5.14), infinitive verb is analyzed as secondary object as marked with dative marker and coming with a ditransitive verb.

(5.15) a.

He	hates	to	fly.
PRON	V		INF V
Nom.3.sg.M			
SUBJ			XCOMP

b. وہ اڑنے سے نفرت کرتا ہے۔

woh	[urney	se]	nafrat	kerta	hey
PRON	V	P	N	V	AUX
Nom.3.sg.M	inf		Nom.3.sg.F	3.sg.M	
SUBJ	OBL		OBJ		

(5.16) a.

He	fears	to	fly.
PRON	V	INF	V
Nom.3.sg.M4			
SUBJ		XCOMP	

b. وہ اڑنے سے ڈرتا ہے۔

woh	urney	se	derta	he.
PRON	V	P	V	AUX
Nom.3.sg	inf		3.sg.M	
SUBJ	OBL			

The above mentioned examples show similar transformations as we have seen in the previous section for objects. This evidence also shows a similarity between XCOMP and OBJ as we have decided to handle XCOMP as OBJ as mentioned in the start of this section.

The rules to implement the mentioned transformations are as follows.

Rule - 9.

#XCOMP_WITH_GEN(arg1, arg2): PRED = arg1, PRED.GF = {<SUBJ,OBJ>},OBJ.PRED = arg2, INSERT(XCOMP,XCOMP_OBJ_GEN), XCOMP_WITH_GEN_FLAG = {TRUE};

Corresponding Insert Rule

XCOMP_OBJ_GEN
[
(t:: XCOMP_WITH_GEN_FLAG) =c {TRUE}) --> OBJ.SPEC.GEN;
]

Rule - 10.

#XCOMP_DITRANS(arg1, arg2): PRED = arg1, PRED.GF = <SUBJ,OBJ,OBJ2>, OBJ.PRED = arg2, INSERT(XCOMP, XCOMP_OBJ2), XCOMP_DITRANS_FLAG = {TRUE};

Corresponding Insert Rule

XCOMP_OBJ2
[
(t::XCOMP_DITRANS_FLAG =c {TRUE}) --> OBJ2;
]

Rule - 11.

#XCOMP_WITH_OBL(arg1, arg2,arg3,arg4): PRED = arg1, PRED.GF = arg4, OBJ.PRED = arg2, OBL.PRED=arg3, OBL.PRED.GF = <OBJ>,INSERT(XCOMP,XCOMP_OBJ_OBL), XCOMP_WITH_OBL_FLAG = {TRUE};

Corresponding Insert Rule

XCOMP_OBJ_OBL
[
(t:: XCOMP_WITH_OBL_FLAG) =c {TRUE}) --> OBL.OBJ;
]

Rule - 12.

#XCOMP_TO_OBL(arg1, arg2,arg3): PRED = arg1, PRED.GF = arg3, OBL.PRED = arg2, OBL.GF = <OBJ>, INSERT(XCOMP, XCOMP_OBLOBJ), XCOMP_OBLOBJ_FLAG = {TRUE};

Corresponding Insert Rule

```
XCOMP_OBLOBJ
[
        (XCOMP_OBLOBJ _FLAG =c {TRUE}) --> OBL.OBJ;
]
```

All verbs have XCOMP subcatagorization are listed in Appendix A.6 along with their respective rules and translations.

5.5 OBL/Adjunct Insertion

In this case verb is translated into a verb in Urdu and an additional prepositional phrase, adding some meanings or specifying manner of the verb. This transformation is categorized as conflatational divergence in Dorr (1994).

In the following example, verb 'clutch' is translated into Urdu verb 'pekerna' and prepositional phrase 'zor se'.

(5.17) a.

He	clutched	the	book.
PRON Nom.3.sg.M	V	ART	N
SUBJ		OBJ	

b. اس نے کتاب زور سے پکڑی۔

[us ne]	kitab	zor se	pekri.
PRON CM Erg.3.sg.M	N Nom.3.sg.F	N P	V 3.sg.F
SUBJ	OBJ	ADJUNCT	

In the above mentioned example, prepositional phrase is analyzed as adjunct. In some cases the prepositional phrase is oblique. Following is the example.

(5.18) a.

He	risked	his	life.
PRON	V	GEN_PRO	N
Nom.3.sg.M			
SUBJ			OBJ

b. اس نے اپنی زندگی خطرے میں ڈالی۔

[us ne]	apni	zindgi	khetrey	men	dali.
PRON CM	PRO	N	N	P	V
Erg.3.sg.M		Nom.3.sg.F			3.sg.F
SUBJ	OBJ		OBL		

The rules to implement above the mentioned transformations are as follows.

Rule - 13.

#ADD_ADJUNCT_PP(arg1, arg2,arg3): PRED = arg1, PRED.GF = PRED.GF,ADJUNCT.PRED = arg2,ADJUNCT.PRED.GF=<OBJ>,ADJUNCT.OBJ.PRED = arg3;

Rule - 14.

#ADD_OBL_PP (arg1, arg2, arg3, arg4): PRED = arg1, PRED.GF = arg2, OBL.PRED = arg3, OBL.PRED.GF=<OBJ>, OBL.OBJ.PRED = arg4;

Verbs following the Rule – 13 are as follows.

Table 5.3: ADJUNCT Insertion Verb List

Verb	Translation	Verb	Translation
clutch	زور سے پکڑنا	dump	مٹی میں دبانا
frame	چوکھٹے میں جڑنا	inherit	وراثت میں ملنا
launch	پانی میں اتارنا	nod	اثبات میں ہلانا
observe	غور سے دیکھنا	retain	لمبا عرصہ تک رکھنا
screen	بڑا پردہ پر لگانا	speed	تیزرفتاری سے جانا
spin	تیزی سے گھمانا	spin	تیزی سے گھومنا
trap	جال میں پھانسنا	whisper	آہستہ سے کہنا

Verbs following the Rule – 14 are as follows.

Table 5.4: OBL Insertion Verb List

Verb	Translation	Verb	Translation
document	تحریر میں لانا	evolve	وجود میں لانا
hire	کرایہ پر لینا	market	بازار میں لانا
rent	کرایہ پر لینا	risk	خطرہ میں ڈالنا
sort	ترتیب سے رکھنا		

5.6 Ditransitive Conversion

5.6.1 Multiple Objects Construction

As we have discussed in Section 1.3.1.4, some verbs can have two noun phrases as objects, referred to as OBJ and OBJtheta. The following shows an example with the verb 'give'.

(5.19) a.

She	gave	him	a	book
PRON	V	PRON	ART	N
Nom.3.sg.F		Acc.3.sg.M		
SUBJ		OBJ		OBJtheta

b. اس نے اسے کتاب دی۔

[us ne]	usey	kitab	di.
PRON CM	PRO	N	V
Erg.3.sg	Dat.3.sg	Nom.3.sg.F	3.sg.F
SUBJ	OBJtheta	OBJ	

As we can see in the above example, the analysis for OBJ and OBJtheta are different in English sentence and its Urdu translation. The following transformation rule is written to handle this difference.

Rule - 15.

#DITRANS_SWAP(arg1): PRED = arg1, PRED.GF = PRED.GF ,INSERT(OBJ, OBJ_OBJTHETA), INSERT(OBJTHETA, OBJTHETA_OBJ), DITRANS_ALT_FLAG) = {TRUE};

Corresponding Insert Rules

```
OBJ_OBJTHETA
[
        (t:: DITRANS_ALT_FLAG) =c {TRUE}) --> OBJtheta;
]
OBJTHETA_OBJ
[
        (t:: DITRANS_ALT_FLAG) =c {TRUE}) --> OBJ;
]
```

Following is another example of a ditransitive verb.

(5.20) a.

She	cooked	him	food
PRON	V	PRON	N
Nom.3.sg.F		Acc.3.sg.M	
SUBJ		OBJ	OBJtheta

b. اس نے اس کے لئے کھانا پکایا۔

[us ne]	[us	ke	liye]	khana	pekaya.
PRON CM	PRO	CM	P	N	V
Erg.3.sg	3.sg			Nom.3.sg.M	3.sg.M
SUBJ	OBL			OBJ	

In the above example, English sentence has ditransitive construction same as in (5.19)a. but semantic relation of OBJ is different in both sentences, in (5.19)a. OBJ is considered as GOAL whereas in (5.20)a. OBJ is considered BENEFICIARY. The meaning of (5.20) a. cannot be conveyed using ditransitive frame in Urdu, so we need to change it into <SUBJ,OBJ,OBL> in Urdu as shown in (5.20) b. A transformation rule is required to deal with ditransitive frame of such verbs. This transformation maps the OBJ of English sentence to OBL in Urdu. Following is the rule for this transformation.

Rule - 16.

#BENF_ALT(arg1): PRED = arg1, PRED.GF = {<SUBJ,OBJ,OBL>}, INSERT(OBJ,OBJ_OBL_GEN,NULL), OBL.PRED='لیے', BENF_ALT _FLAG) =c {TRUE};

Corresponding Insert Rule

```
OBJ_OBL_GEN
[
        (t:: BENF_ALT _FLAG) =c {TRUE}) --> OBL.SPEC.GEN.GENOBJ;
]
```

5.6.2 Oblique Construction

The dative construction mentioned in the above section in example (5.19) has an alternation in which the same meaning can be conveyed with SUBJ, OBJ, OBL functions where OBJ of the above mentioned construction is changed into OBL and OBJtheta into OBJ. In Urdu there is only one way of expressing ditransitive verbs, that is SUBJ, OBJ, OBJtheta where OBJtheta is marked with dative case marker. Following example shows the alternation of sentence mentioned in example (5.19) and its translation.

(5.21) a.

She	gave	a	book	to	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F					Acc.3.sg.m
SUBJ			OBJ		OBL

b. اس نے اسے کتاب دی۔

[us ne]	usey	kitab	di.
PRON CM	PRO	N	V
Erg.3.sg	Dat.3.sg	Nom.3.sg.F	3.sg.F
SUBJ	OBJtheta	OBJ	

A transformation is required to deal with prepositional frame of these verbs. This transformation maps OBL of English to OBJtheta of Urdu. Following is the rule for this transformation.

Rule - 17.

#DITRANS_ALT(arg1): PRED = arg1, PRED.GF = {<SUBJ, OBJ, OBJtheta>},INSERT(to_p, NULL), INSERT(OBL, OBL_OBJtheta, NULL), DITRANS_ALT_FLAG) = {TRUE};

Corresponding Insert Rule

OBL_OBJtheta
[
(t:: DITRANS_ALT_FLAG) =c {TRUE}) --> OBJtheta;
]

5.7 Other Transformations

In this section individual verbs are discussed which do not fall in the above mentioned categories.

In the following example different sentences with the verb 'share' are shown.

(5.22) a.

She	shared	the	idea	with	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F		Acc.3.sg			Acc.3.sg.M
SUBJ		OBJ		OBL	

b. اس نے اسے خیال بتایا۔

[us ne]	usey	khiyal	betaya
PRON CM	PRON	PRON	V
Erg.3.sg	Dat.3.sg.M	3.sg.M	3.sg.M
SUBJ	OBJtheta	OBJ	

(5.23) a.

She	shared	the	cake	with	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F		Acc.3.sg			Acc.3.sg.M
SUBJ		OBJ		OBL	

b. اس نے کیک اس کے ساتھ مل کر کھایا۔

[us ne]	kek	us ke sath	mil ker	khaya
PRON CM	N	PRON P		V
Erg.3.sg	3.sg	3.sg.M		3.sg.M
SUBJ	OBJ	OBL	ADJUNCT	

(5.24) a.

She	shared	the	room	with	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F		Acc.3.sg			Acc.3.sg.M
SUBJ		OBJ		OBL	

b. اس نے کمرہ اس کے ساتھ مل کر استعمال کیا۔

[us ne]	kemra	us ke sath	mil ker istmal kia	
PRON CM	N	PRON P	N	V
Erg.3.sg	3.sg	3.sg.M		3.sg.M
SUBJ	OBJ	OBL	ADJUNCT	

As we can see in the above mentioned sentences, the verb ‘share’ is difficult to translate precisely in Urdu. In sentence (5.22) the object of the sentence is abstract in nature and the verb ‘share’ is giving the meaning of telling somebody about the abstract entity, e.g. thought, idea. So the verb is translated as ‘betana’ (to tell) in Urdu. For sense conveyed in (5.23) and (5.24), there is no straight translation in Urdu; translation is varying according to object being shared. To have a workable solution, verb ‘share’ is transliterated and a verbal noun construction is made with verb ‘kerna’. Sentences are translated as below.

(5.25) a.

She	shared	the	idea	with	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F		Acc.3.sg			Acc.3.sg.M
SUBJ		OBJ		OBL	

b. اس نے خیال اس کے ساتھ شیئر کیا۔

[us ne]	khial	us ke sath	share	kiya
PRON CM	N	PRON P		V
Erg.3.sg	3.sg	3.sg.M		3.sg.M
SUBJ	OBJ	OBL		

c.

She	shared	the	cake	with	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F		Acc.3.sg			Acc.3.sg.M
SUBJ		OBJ		OBL	

d. اس نے کیک اس کے ساتھ شیئر کیا۔

[us ne]	kek	us ke sath	share	kiya
PRON CM	N	PRON P		V
Erg.3.sg	3.sg	3.sg.M		3.sg.M
SUBJ	OBJ	OBL		

e.

She	shared	the	room	with	him
PRON	V	ART	N	P	PRON
Nom.3.sg.F		Acc.3.sg			Acc.3.sg.M
SUBJ		OBJ		OBL	

f. اس نے کمرہ اس کے ساتھ شیئر کیا۔

[us ne]	kemra	us ke sath	share	kia
PRON CM	N	PRON P		V
Erg.3.sg	3.sg	3.sg.M		3.sg.M
SUBJ	OBJ	OBL		

A common strategy for introducing new verbs in Urdu is adding ‘kerna (do)’ or ‘hona (be)’ to a new word to produce a complex predicate. This transliteration strategy is used in the MT system when no simple translation of a verb exists or when different translations are required for different senses which are hard to disambiguate. In such cases the verb is transliterated to avoid sense disambiguation. Code switching (use of foreign language words in a language) is another factor that is considered when using this strategy, i.e., when an English language word is judged by native speakers to be of common use in the Urdu language, instead of translating it, it is transliterated. Examples of such verbs can found in Appendix A.1 which contains verbal noun translation rules. Some examples are given below:

Table 5.4: Sample Transliterated Verbs

Verb	Translation	Verb	Translation
Bounce	باؤنس ہونا	Cancel	کینسل کرنا
Cast	کاسٹ کرنا	Set	سیٹ کرنا

In the following example, there is no direct construction to convey the meaning of 'overlook' and object in the English sentence is converted into SUBJ in Urdu whereas SUBJ in English is OBL in the Urdu sentence.

(5.26) a.

The room	overlooks	the	ocean.
ART N	V	ART	N
Nom.3.sg		Acc.3.sg	
SUBJ		OBJ	

c. کمرے سے سمندر نظر آتا ہے۔

[kemrey se]	semender	[nezar	aata	he]
N P	N	N	V	AUX
3.sg	3.sg		3.sg.M	
OBL	SUBJ			

6 Discussion

The rules listed in the above section may co-occur in the same verb. The rule that most frequently co-occurs with other rules is the verbal noun rule (section 5.1). Words translated into verbal nouns exhibit behavior similar to that of simple verbs and during this process other rules may also be applied simultaneously. Interaction between other rules also occurs, and this is illustrated using the examples that follow.

(6.1)	a.	He	ordered	him	to	leave	
		PRON Nom.3.sg.M SUBJ	V	PRON Acc.3.sg.M OBJ	INF XCOMP	V	
	b.	اس نے اسے جانے کا حکم دیا۔					
		[Us ne]	usey	[janey ka	hukam]		diya.
		PRON CM Erg.3.sg.F SUBJ	PRON Dat.3.sg OBJtheta	V CM	N Nom.3.sg.M OBJ		V:GIVE 3.sg.M

In the above example two rules are applied, XCOMP to OBJ Conversion and Object Insertion. The order of these rules is important. In the example, we can see that the XCOMP to OBJ rule is applied first and then the Transitive Rule for Object Insertion is applied, Rule - 3.

In example 6.2, the verb ‘remind’ is translated to the verbal noun ‘یاد دلانا’ by applying Rule 1. The XCOMP is treated as a noun phrase and then a Ditransitive Transformation is applied to convert the English OBJ an Urdu OBJtheta.

(6.2)	a.	She	reminded	me	[to	buy	milk]
		PRON Nom.3.sg.F	V	PRON Acc.1.sg	INF	V	N

SUBJ OBJ XCOMP

b. اس نے مجھے دودھ خریدنا یاد دلایا۔

[Us ne]	mujhe	[dodh khereedna]	yad	dilaya.
PRON CM	PRON	N V	N	V
Erg.3.sg.F	Dat.3.sg	Nom.3.sg.M	Nom.3.sg.F	3.sg.M
SUBJ	OBJtheta		OBJ	

By analyzing examples, including the ones described above, an order for rule application can be deduced. The XCOMP Rule is applied first. The XCOMP is converted to an OBJ or an OBJtheta if the English sentence already contains an OBJ. Then the remaining rules are applied. If the earlier mentioned proposal for excluding the COMP and the XCOMP in grammar analysis is eventually implemented, there will be no need of the XCOMP Conversion Rules described (Section 5.4) and this rule application order.

The order of rule application of other rules in conjunction with the Verbal Noun Rule is insignificant.

Theoretically speaking, rules should be applied in the order described above, but the current MT system only allows the application of one rule for each translation. Due to this, when multiple rules apply to a single translation, a single rule has to be produced that contains all the required rules, and rule ordering is not actually reflected in the MT system.

In some cases rules have to be repeated redundantly due to limitations of the system. This problem can be illustrated using the example of the verb ‘concentrate’. Consider the following use of Rule 2 presented in section 5.1.1.

```
concentrate_v
[
        (PRED.GF =c {<SUBJ>}) --> #TRANS('de', 'tewajo',
<SUBJ,OBJ>);

        (PRED.GF =c {<SUBJ,OBL>}) --> #TRANS('de', 'tewajo',
<SUBJ,OBJ,OBL>);
]
```

As can be seen above, the rule needs to be repeated for two different subcategorization frames. Moreover, if some new subcategorization is found with the same transformation, the rule writer has to add a new rule to deal with the newly found subcategorization. To better realize the rule, the syntax should have the provision to add and delete grammatical functions. This will provide the flexibility to state that for this particular set of subcategorizations i.e. for all the intransitive ones in the above mentioned example, add an object to the Urdu structure.

We have seen in many rules i.e. Rules 3-7, that a flag is used to identify the proper place to use the overriding specialized rule. This method is not very reliable as it is completely the rule writer's responsibility to make sure that a flag is used and is uniquely identifiable. A better solution will be to assign a name to the required structure and then using that name to apply the rule.

This work mainly discusses lexical-semantic divergences except in Section 5.6 of ditransitive verbs. These rules fall under the category of syntactic divergence but are discussed here because there are two classes of ditransitive verbs, having either a beneficiary or a goal as the object. Lexical semantic divergences presented by Dorr (1994) are discussed in section 2.4. Some transformations discussed in the work are similar to Dorr's divergences, such as OBJ / OBL which is called structural divergence in Dorr's work. The other one is OBL / ADJUNCT insertion which is called conflatation in Dorr's work. Other transformations discussed in the work are not addressed in Dorr's work. The phenomenon of Object Insertion which is discussed in Section 5.2 is observed in Urdu very frequently. It does not fit into any category mentioned by Dorr. Another new found transformation is the analysis of the XCOMP in English as a noun phrase in

Urdu. There are transformations mentioned in Dorr's work which are not listed in this work.

An attempt to establish correspondence of rules in the study with Levin (1993) verb classes was made but no significant correspondence between these two classes was identified. There were few similarities found in the MT rules and Levin (1993) verb classes. Verbs such as the GIVE verbs of Levin (1993) were mostly translated as the ditransitive verb 'dena' in Urdu or followed the Object Insertion rule with secondary object construction as was described in Section 5.2.2.1. Some of the verbs were translated into the dative verb 'dena' with manner explicitly added as an ADJUNCT/ OBL. For example, the verb 'rent' which is a member of the GIVE class is translated into 'karaye per dena'. Such correspondence does not hold when we look at the verbs on which MT rules are applied. Each rule has verbs, which are member of different classes.

This study may be useful for other languages where similar phenomena occur, especially South Asian languages which are linguistically similar to Urdu. Phenomena such as complex predication and infinitive verbs acting as nouns are common in many South Asian languages. These phenomena do not exist in other languages, and transformation rules are needed for translation between languages where these phenomena occur and other languages. The work presented in this thesis will aid in the development of such rules.

References

Alsina, Alex, Mohanan,T and Mohanan,K.P. (2005), How to get rid of the COMP. Proceedings of the LFG05 conference.

Apresian, J., Boguslavskij, I., Iomdin, L., Lazurskij, A., Sannikov, V. and Tsinman, L. (1992), ETAP-2: The linguistics of a Machine Translation system. *META*, XXXVII(4):97–112.

Arnold, D.J. (1986), Eurotra: A European perspective on MT. *Proceedings of the IEEE*, 74(7): 979–992.

Arnold, D.J. and Tombe, L. (1987), Basic theory and methodology in Eurotra. In Sergei Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues*, pages 114–135. Cambridge University Press, Cambridge.

Arnold D.J. and Sadler, L.G. (1990), The theoretical basis of MiMo. *Machine Translation*, 5:195–222.

Arnold D., Balkan L., Humphreys R. Lee, Meijer S., and Sadler, L. (1993), Machine Translation: An Introductory Guide, NCC Blackwell, Manchester, Oxford.

Bennett, W. S. and Slocum, J. (1988), METAL: the LRC Machine Translation system. In J. Slocum, editor, *Machine Translation Systems*, pages 111–140. CUP, Cambridge.

Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin (1990), A statistical approach to machine translation. Computational Linguistics, 16(2):76-85.

Butt M. (1995), The Structure of Complex Predicates in Urdu. CSLI Publications.

Butt M., King T.H., Nifio M.E., and F. Segond. (1999), A Grammar Writer's Cookbook. CSLI Publications, Stanford, CA.

Copeland, C., Durand, J., Krauwer, S. and Maegaard, B. (1991a) *The Eurotra Formal Specifications*, volume 2 of *Studies in Machine Translation and Natural Language Processing*. Office for Official Publications of the Commission of the European Community,Luxembourg.

Copeland, C., Durand, J., Krauwer, S. and Maegaard, B. (1991b) *The Eurotra Linguistic Specifications*, volume 1 of *Studies in Machine Translation and Natural Language Processing*. Office for Official Publications of the Commission of the European Community, Luxembourg.

Dalrymple, Mary E. (2001), Lexical functional grammar (Syntax and semantics, vol. 34). NY: Academic Press.

Dorr. B. J. (1994), "Machine Translation Divergences: A Formal Description and Proposed Solution", *ACL Vol. 20, No. 4, pp. 597-631.*

Estival, D., Ballim, A., Russell, G. and Warwick, S. (1990), A syntax and semantics for feature-structure transfer. In *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, pages 131–143, Linguistics Research Centre, University of Texas at Austin, USA.

Forst, M. (2006), COMP in (parallel) Grammar Writing. Proceedings of the LFG06 conference.

Goodman, K. (1989), *Machine Translation*, volume 4. numbers 1 and 2 : Special Issues on Knowledge Based MT, Parts I and II.

Goodman, K. and Nirenburg, S. (1991), editors. *The KBMT Project: A Case Study in Knowledge Based Machine Translation*. Morgan Kaufmann, San Mateo, California.

Gupta D. and Chatterjee N. (2001), Study of Divergence for Example Based English-Hindi Machine Translation in *'Proceedings of Symposium on Translation Support Systems'* 15 February – 17 February, 2001.

Halvorsen, P. K. and Kaplan, R. M. (1988), Projections and semantic description in LFG, *in* '*Proceedings of the International Conference on Fifth Generation Computer Systems* (FGCS '88), Tokyo, 28 November - 2 December 1988', Institute for New Generation Computer Technology, Tokyo, pp. 1116-1122.
Also appeared in: Dalrymple, M., Kaplan, R. M., Maxwell, J. T. and Zaenen, A., eds (1995), *Formal issues in Lexical-Functional Grammar*, Center for the Study of Language and Information, Stanford, CA, pp. 293-309.

Hutchins, W. J. and Somers, H. L. (1992), *An introduction to machine translation*, Academic Press, London.

Isabelle, P. (1987), Machine Translation at the TAUM group. In M. King, editor, *Machine Translation Today: The State of the Art, Proceedings of the Third Lugano Tutorial, 1984*, pages 247–277. Edinburgh University Press, Edinburgh.

Jurafsky and J. H. Martin (2000), Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, Upper Saddle River,NJ,.

Kaplan, R. M. (1989), The formal architecture of Lexical-Functional Grammar, *in* C. R. Huang and K. J. Chen, eds, 'Proceedings of the Republic of China Computational Linguistics Conference (ROCLING II), Taipei, 1989', Academia Sinica, Taipei, pp. 1-18. Also appeared in: Dalrymple, M., Kaplan, R. M., Maxwell, J. T. & Zaenen, A., eds (1995), *Formal issues in Lexical-Functional Grammar*, Center for the Study of Language and Information, Stanford, CA, pp. 7-27.

Kaplan, R. M., Netter, K., Wedekind, J. & Zaenen, A. (1989), Translation by structural correspondences, *in* 'Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics, UMIST, Manchester, 10-12 April 1989', Association for Computational Linguistics, New Brunswick, NJ, pp. 272-281. Also appeared in: Dalrymple, M., Kaplan, R. M., Maxwell, J. T. & Zaenen, A., eds (1995), *Formal issues in Lexical-Functional Grammar*, Center for the Study of Language and Information, Stanford, CA, pp. 311-319.

Kay, M. (1984), Functional Unification Grammar: a formalism for machine translation, *in* 'Proceedings of the 10th International Conference on Computational Linguistics (COLING '84), Stanford, CA, 2-6 July 1984', Stanford University, Stanford, pp. 75-78.

Kroeger, Paul R. (2005), Analyzing grammar: an introduction. New York: Cambridge University Press

Kudo, I. and Nomura, H. (1986), Lexical-Functional Transfer: a transfer framework in a machine translation system based on LFG, *in* 'Proceedings of the 11th International Conference on Computational Linguistics (COLING '86), Bonn, 25-29 August 1986', University of Bonn, Bonn, pp. 112-114.

Landsbergen, J.S.P. (1987a), Isomorphic grammars and their use in the Rosetta translation system. In Margaret King, editor, *Machine Translation Today: The State of the Art, Proceedings of the Third Lugano Tutorial, 1984*, pages 351–372. Edinburgh University Press, Edinburgh.

Landsbergen, J.S.P. (1987b), Montague grammar and Machine Translation. In Peter Whitelock, Mary McGeeWood, Harold L. Somers, Rod L. Johnson, and Paul Bennett, editors, *Linguistic Theory and Computer Applications*, pages 113–147. Academic Press, London.

Longman Dictionary of Contemporary English (1978), Longman Group Limited, London.

Maas, H. (1987) The MT system SUSY. In Margaret King, editor, *Machine Translation Today: The State of the Art, Proceedings of the Third Lugano Tutorial, 1984*, pages 209–246. Edinburgh University Press, Edinburgh.

McCord, M.C. (1989), Design of LMT: A Prolog-based Machine Translation system. *Computational Linguistics*, 15(1):33–52.

Mohanan T. (1994), *Argument Structure in Hindi*. Stanford, California: CSLI Publications.

Nagao, M., Tsujii, J. and Nakamura, J. (1986), Machine Translation from Japanese into English. *Proceedings of the IEEE*, 74(7):993–1012.

Noord, G. van, Dorrepaal, J., Eijk, P. van der, Florenza, M. and Tombe, L. des. (1990), The MiMo2 research system. In *Third International Conference on Theoretical and Methodological Issues in Machine Transation 11-13 June 1990*, pages 213–124, Linguistics Research Center, Austin, Texas.

Sharp, R. (1988) CAT-2 – implementing a formalism for multi-lingual MT. In *Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 76–87, Carnegie Mellon University, Centre for Machine Translation, Pittsburgh, USA.

Slocum, J., Bennett, W.S., Bear, J., Morgan, M. and Root, R. (1987), Metal: The LRC Machine Translation system. In M. King, editor, *Machine Translation Today*, pages 319–350. Edinburgh University Press, Edinburgh.

Talmy, L. (1985),'Lexicalization patterns: semantic structure in lexical forms' in Shopen ed. Language Typology and Syntactic Description III: Grammatical categories and the lexicon. CUP. 57-149

Vauquois, B. and Boitet, C. (1985), Automated translation at Grenoble University. *Computational Linguistics*, 11(1):28–36.

WordNet 2.0, http://wordnet.princeton.edu/cgi-bin/webwn

Appendix A: List of Verbs

Appendix A.1: List of verbs for Verbal Noun Conversion Rule, R-1

Verb	Urdu Translation	Verb	Urdu Translation
absorb	جزب کرنا	initiate	شامل کرنا
abuse	تباہ کرنا	inject	داخل کرنا
accelerate	تیزہونا	injure	زخمی کرنا
accept	قبول کرنا	install	نصب کرنا
accumulate	اکھٹا کرنا	introduce	متعارف کرنا
achieve	حاصل کرنا	invent	ایجادکرنا
acknowledge	تسلیم کرنا	invite	مدعوکرنا
acquire	حاصل کرنا	invoke	عائد کرنا
activate	متحرک کرنا	isolate	حاصل کرنا
adapt	تشکیل کرنا	issue	جاری کرنا
Add	اضافہ کرنا	justify	جسٹفائی کرنا
address	مخاطب کرنا	Kick	ترک کرنا
adjust	مانوس ہونا	Kill	ہلاک ہونا
admire	پسند کرنا	Lean	کھڑا کرنا
affect	متاثر کرنا	learn	معلوم ہونا
afford	پیش کرنا	level	ہموارکرنا
agree	قبول کرنا	light	روشن کرنا
alert	خبردار کرنا	Like	پسندہونا
alter	کھلا کرنا	limit	محدود کرنا

appear	نمودارہونا	Line	پر کرنا
approve	منظورکرنا	List	درج کرنا
arise	پیداہونا	Live	زندہ رہنا
arouse	بیدارکرنا	locate	معلوم کرنا
arrest	گرفتارکرنا	Lock	بند کرنا
arrive	پیداہونا	lodge	دائر کرنا
Ask	مول لینا	Lose	کم کرنا
assign	مختص کرنا	love	پسند ہونا
associate	منسوب کرنا	lower	کم کرنا
assume	اختیارکرنا	maintain	برقراررکھنا
assure	یقینی بنانا	manipulate	استعمال کرنا
attach	مانوس ہونا	mark	مارک کرنا
attain	حاصل کرنا	melt	ختم ہونا
attend	شریک ہونا	merge	ضم ہونا
attract	حاصل کرنا	miss	یاد کرنا
attribute	منسوب کرنا	model	ماڈل کرنا
back	ثابت کرنا	motivate	راغب کرنا
balance	متوازن کرنا	mount	منعقد کرنا
bang	بندہونا	move	منتقل کرنا
bear	برداشت کرنا	murder	قتل کرنا
begin	شروع ہونا	neglect	نظراندازکرنا
behave	پیش آنا	nominate	منتخب کرنا
bind	پابند کرنا	obscure	غیرواضح کرنا

Bite	بند رکھنا	obtain	رائج ہونا
block	بند کرنا	occupy	مشغول رکھنا
bother	متفکر ہونا	offer	پیش کرنا
bounce	باؤنس ہونا	offset	متوازن کرنا
burst	بھرا ہونا	omit	نظرانداز کرنا
Call	ٹیلیفون کرنا	open	شروع ہونا
calm	کم کرنا	originate	شروع ہونا
cancel	کینسل ہونا	overlook	نظرانداز کرنا
capture	محفوظ کرنا	pack	تیار کرنا
carve	کندہ کرنا	paint	پیش کرنا
Cast	کاسٹ کرنا	park	پارک کرنا
catch	نظر آنا	part	الگ کرنا
cause	پیدا کرنا	participate	شریک ہونا
challenge	چیلنج کرنا	pass	پاس ہونا
check	چیک کرنا	penetrate	داخل ہونا
Cite	طلب کرنا	persist	برقرار رہنا
claim	قبول کرنا	persuade	قائل کرنا
clarify	واضح کرنا	pick	صاف کرنا
clean	صاف کرنا	position	متعین کرنا
clear	صاف کرنا	possess	متاثر کرنا
collapse	ختم ہونا	postpone	ملتوی کرنا
collect	جمع کرنا	prepare	تیار کرنا
commence	شروع ہونا	prescribe	تجویز کرنا

compile	مرتب کرنا	present	پیش کرنا
compose	تشکیل دینا	preserve	محفوظ کرنا
concede	تسلیم کرنا	press	استری کرنا
concentrate	مجتمع کرنا	presume	فرض کرنا
concern	پریشان کرنا	pretend	ظاہر کرنا
conclude	اخذ کرنا	process	تیارکرنا
conduct	منتقل کرنا	proclaim	قراردینا
confine	محدود کرنا	progress	بہترہونا
connect	کنیکٹ کرنا	prohibit	منع کرنا
constitute	قائم کرنا	pronounce	قراردینا
consume	استعمال کرنا	protect	محفوظ رکھنا
continue	برقراررکھنا	prove	ثابت کرنا
convert	داخل کرنا	publish	شائع کرنا
convict	مجرم ٹھہرانا	pursue	اختیارکرنا
convince	قائل کرنا	push	مجبورکرنا
correct	ٹھیک کرنا	raise	اونچا کرنا
cover	ط کرنا	reach	قائل کرنا
crash	کریش ہونا	react	متاثرہونا
cross	پارکرنا	realise	حقیقت بنانا
declare	قراردینا	rebuild	بہتربنانا
decline	کم ہونا	recall	یادکرنا
dedicate	وقف کرنا	receive	قبول کرنا
defeat	ردکرنا	recognize	تسلیم کرنا

define	بیان کرنا	record	ریکارڈ کرنا
delay	ملتوی کرنا	recover	شفایاب ہونا
deliver	پیدا کرنا	recruit	بھرتی کرنا
demonstrate	واضح کرنا	reduce	کم کرنا
depart	روانہ ہونا	refuse	مسترد کرنا
depend	منحصر ہونا	regain	بحال کرنا
deprive	محروم کرنا	reinforce	مضبوط کرنا
derive	ماخوذ کرنا	reject	مسترد کرنا
detect	محسوس کرنا	relax	خاموش ہونا
determine	معین کرنا	release	فارغ کرنا
develop	منظم بنانا	relieve	کم کرنا
devise	ایجاد کرنا	remark	تبصرہ کرنا
devote	وقف کرنا	remember	یاد رکھنا
differ	مختلف ہونا	remind	یاد دلانا
differentiate	تفریق کرنا	render	فراہم کرنا
diminish	کم ہونا	repay	واپس کرنا
Dip	نیچے ہونا	reserve	مخصوص کرنا
disagree	مختلف ہونا	resolve	علیحدہ کرنا
disappear	ختم ہونا	restore	بحال کرنا
discharge	فارغ کرنا	restrict	محدود کرنا
dismiss	مسترد کرنا	retain	برقرار رکھنا
display	ظاہر کرنا	retire	ریٹائر ہونا
dissolve	حل ہونا	reveal	ظاہر کرنا

distinguish	تمیز کرنا	reverse	پیچھے کرنا
distribute	تقسیم کرنا	revive	زندہ کرنا
disturb	پریشان کرنا	Rid	پاک کرنا
divert	منتقل کرنا	Rise	بلند ہونا
divide	تقسیم کرنا	round	گول کرنا
dominate	حاوی ہونا	Sail	روانہ ہونا
draft	بھرتی کرنا	satisfy	مطمئن کرنا
draw	اخذ کرنا	scan	سکین کرنا
drop	کم کرنا	seal	ہوابند کرنا
Dry	خشک کرنا	search	استعمال کرنا
ease	کم کرنا	secure	حاصل کرنا
effect	متاثر کرنا	seize	ضبط کرنا
eliminate	ختم کرنا	select	منتخب کرنا
embark	سوارہونا	sense	محسوس کرنا
emerge	نمودارہونا	serve	پیش کرنا
emphasize	نمایاں کرنا	Set	سیٹ کرنا
employ	رائج کرنا	settle	ختم کرنا
empty	خالی کرنا	shape	تشکیل کرنا
enclose	ملفوف کرنا	share	شیئر کرنا
enforce	مسلط کرنا	shed	کم کرنا
engage	مصروف ہونا	shift	منتقل ہونا
enhance	بہتر کرنا	shut	بند ہونا
enter	داخل ہونا	Sink	نیچے ہونا

equip	آراستہ کرنا	situate	پیش کرنا
erect	قائم کرنا	slow	آہستہ ہونا
escape	فرارہونا	smash	ختم کرنا
exchange	تبدیل کرنا	smooth	ہموارکرنا
exercise	استعمال کرنا	solve	حل کرنا
exhibit	ظاہرکرنا	Sort	علیحدہ کرنا
exist	موجودہونا	spare	بخش دینا
export	برآمد کرنا	specify	واضح کرنا
expose	بےنقاب کرنا	spend	خرچ کرنا
extend	لمباہونا	split	تقسیم ہونا
extract	حاصل کرنا	spoil	خراب کرنا
Fail	ناکام ہونا	stand	کھڑاہونا
Fear	خوفزدہ ہونا	start	شروع ہونا
feature	نمایاں ہونا	strengthen	مضبوط کرنا
Feel	محسوس ہونا	submit	پیش کرنا
figure	شمارہونا	succeed	کامیاب ہونا
file	درج کرنا	suffer	مبتلاہونا
Find	معلوم ہونا	suggest	ظاہرکرنا
finish	ختم ہونا	summon	طلب کرنا
Fire	فائرکرنا	supply	مہیاکرنا
Fit	پوراہونا	suppose	فرض کرنا
Fix	مقررکرنا	suppress	کمزورکرنا
flick	بند کرنا	surprise	حیران کرنا

flush	فلش کرنا	suspend	معطل کرنا
focus	فوکس کرنا	sustain	قائم رکھنا
forbid	منع کرنا	sweep	صاف کرنا
forgive	معاف کرنا	swonاs	ظاہر کرنا
form	قائم کرنا	take	قبول کرنا
formulate	پیش کرنا	Tap	حاصل کرنا
found	قائم کرنا	terminate	ختم کرنا
frame	نمایاں کرنا	throw	پیدا کرنا
Free	رہا کرنا	thrust	وارکرنا
freeze	خراب ہونا	tighten	مضبوط کرنا
fulfill	پورا کرنا	tolerate	برداشت کرنا
Gain	حاصل کرنا	transfer	منتقل ہونا
gather	جمع ہونا	transform	تبدیل کرنا
generate	پیدا کرنا	translate	تبدیل کرنا
govern	متعین کرنا	transmit	منتقل ہونا
grant	منظورکرنا	trigger	شروع کرنا
grip	محوکرنا	type	ٹائپ کرنا
heat	گرم کرنا	undermine	کھوکھلا کرنا
highlight	نمایاں کرنا	unite	متحدہونا
Hit	زخمی کرنا	update	اپڈیٹ کرنا
hunt	تلاش کرنا	Use	استعمال کرنا
hurry	مجبورکرنا	vanish	غائب ہونا
ignore	نظراندازکرنا	walk	براہونا

illustrate	واضح کرنا	warm	گرم کرنا
imagine	محسوس کرنا	warn	خبردارکرنا
imply	ثابت کرنا	wash	صاف کرنا
import	درآمد کرنا	waste	ضائع کرنا
impose	مسلط کرنا	watch	خیال رکھنا
impress	متاثرکرنا	weaken	کمزورہونا
improve	بہترہونا	widen	وسیع کرنا
include	شامل کرنا	Win	حاصل کرنا
incorporate	شامل کرنا	wipe	صاف کرنا
indicate	ظاہرکرنا	work	کام کرنا
influence	متاثرکرنا	worry	پریشا ن ہونا
inform	متاثرکرنا	wound	زخمی کرنا

Appendix A.2: List of verbs for Object Insertion for Intransitive Verbs, R-2

Verb	Urdu Translation	Verb	Urdu Translation
advise	مشورہ دینا	kick	ٹھوکر مارنا
analyse	تجزیہ کرنا	knit	بنائی کرنا
appeal	اپیل کرنا	lead	رہنمائی کرنا
arrange	انتظام کرنا	march	مارچ کرنا
attack	حملہ کرنا	marry	شادی کرنا
believe	ایمان رکھنا	move	موقف بدلنا
benefit	فائدہ پہنچنا	object	اعتراض ہونا
bet	شرط لگانا	offer	پیشکش کرنا

bite	چارہ پکڑنا	operate	آپریشن کرنا
boast	شیخی بگھارنا	order	آرڈر دینا
book	بکنگ کروانا	paint	تصویر بنانا
bother	زحمت کرنا	pay	ادائیگی کرنا
celebrate	خوشی منانا	phone	ٹیلیفون کرنا
change	کپڑے بدلنا	plan	منصوبہ بنانا
chase	پیچھا کرنا	plead	بھیک مانگنا
chat	گپ شپ کرنا	pour	بارش ہونا
cheer	حوصلہ افزائی کرنا	practice	مشق کرنا
claim	مطالبہ کرنا	pray	دعا کرنا
clean	صفائی کرنا	promise	وعدہ کرنا
command	حکم دینا	prompt	اشارہ دینا
comment	تبصرہ کرنا	pronounce	اعلان کرنا
communicate	رابطہ کرنا	protest	احتجاج کرنا
compete	مقابلہ کرنا	react	ردعمل ظاہرکرنا
complain	شکایت کرنا	recruit	بھرتی کرنا
compose	نغمہ سازی کرنا	register	اندراج کرنا
concentrate	توجہ دینا	relax	آرام کرنا
conform	پیروی کرنا	reply	جواب دینا
consult	مشورہ کرنا	research	تحقیق کرنا
cook	کھانا پکانا	resign	استعفا دینا
copy	نقل کرنا	rest	آرام کرنا
count	گنتی گننا	ride	سواری کرنا

counter	اختلاف کرنا	ring	فون کرنا
criticize	تنقید کرنا	roar	شورکرنا
cross	آلتی پالتی مارنا	rub	رگڑ لگانا
dare	ہمت کرنا	rule	حکومت کرنا
decide	فیصلہ کرنا	rush	جلدی کرنا
decline	معذرت کرنا	sail	کشتی رانی کرنا
delay	دیر کرنا	score	نمبر بنانا
demonstrate	مظاہرہ کرنا	serve	خدمت کرنا
die	جان دینا	shoot	گولی چلانا
dive	غوطہ خوری کرنا	shrug	کندھا اچکانا
draw	تصویر بنانا	sigh	آہ بھرنا
dream	پلاؤ خیالی پکانا	signal	اشارہ کرنا
dress	لباس پہننا	smoke	سیگریٹ پینا
drink	شراب پینا	sniff	ناک چڑھانا
entertain	تفریح مہیانا	speak	بات کرنا
exercise	ورزش کرنا	steal	چوری کرنا
exhibit	نمائش کرنا	supervise	نگرانی کرنا
exist	گزارہ کرنا	surrender	ہتھیار ڈالنا
explore	دورہ کرنا	swallow	تھوک نگلنا
fish	مچھلی پکڑنا	swear	بدزبانی کرنا
guess	اندازہ لگانا	swing	جھولا جھولنا
hit	حملہ کرنا	talk	بات کرنا
hunt	شکارکرنا	tour	دورہ کرنا

hurry	جلدی کرنا	trade	تجارت کرنا
inherit	وراثت ملنا	translate	ترجمہ کرنا
insist	اصرارکرنا	try	کوشش کرنا
interfere	مداخلت کرنا	visit	سیر کرنا
interpret	ترجمہ کرنا	vote	ووٹ ڈالنا
interrupt	روکاٹ ڈالنا	weave	بنائی کرنا
invest	سرمایہ کاری کرنا	whisper	سرگوشی کرنا
investigate	تفتیش کرنا	work	کام کرنا
judge	اندازہ لگانا	yield	ہارماننا
jump	چھلانگ لگانا		

Appendix A.3: List of verbs for R-3

target	بنا	ہدف	relieve	دے	چھٹی	score	دے	نمبر
dress	پہنا	لباس	command	دے	حکم	answer	دے	جواب
relax	پہنچا	آرام	invite	دے	دعوت	signal	کر	اشارہ
rest	پہنچا	آرام	support	دے	سہارا	manipulate	کر	جوڑتوڑ
benefit	پہنچا	فائدہ	promote	دے	فروغ	ring	کر	فون
harm	پہنچا	نقصان	advise	دے	مشورہ	phone	کر	ٹیلیفون
damage	پہنچا	نقصان	propose	دے	تجویز	exercise	کرا	ورزش
calm	دلا	اطمینان	prefer	دے	ترجیح	feed	کھلا	کھانا
inspire	دلا	جوش	favour	دے	ترجیح	stab	گھونپ	خنجر
hang	دے	پھانسی	formulate	دے	تشکیل	smash	لگا	ضرب
prompt	دے	اشارہ	educate	دے	تعلیم	shoot	مار	گولی
value	دے	اہمیت	trouble	دے	تکلیف	kick	مار	ٹھوکر

curl	دے	بل	wind	دے	چابی	vote	دے	ووٹ

Appendix A.4: List of verbs for R-4

Verb	Urdu Translation	Verb	Urdu Translation
administer	انتظام کرنا	list	فہرست بنانا
advocate	حمایت کرنا	maintain	دیکھ بھال کرنا
aim	نشانہ باندھنا	market	مشہوری کرنا
analyse	تجزیہ کرنا	marry	شادی کروانا
announce	اعلان کرنا	measure	پیمائش کرنا
apologize	معذرت کرنا	miss	موقع گنوانا
appreciate	اندازہ لگانا	model	ماڈلنگ کرنا
arrange	انتظام کرنا	mount	آغازکرنا
assess	اندازہ لگانا	name	نام رکھنا
assist	مدد کرنا	observe	مشاہدہ کرنا
assure	یقین دلانا	offer	پیشکش کرنا
attempt	کوشش کرنا	oppose	مخالفت کرنا
attend	دیکھ بھال کرنا	order	آرڈردینا
back	مدد کرنا	outline	خاکہ کھینچنا
ban	ممانعت کرنا	owe	مقروض ہونا
bet	شرط لگانا	paint	نقشہ کھینچنا
bid	بولی لگانا	pay	ادائیگی کرنا
boast	حامل ہونا	permit	اجازت دینا
book	بکنگ کروانا	picture	تصورکرنا

bother	زحمت کرنا	plan	منصوبہ بنانا
calculate	حساب لگانا	plead	بھیک مانگنا
celebrate	تعریف کرنا	pledge	وعدہ کرنا
chair	صدارت کرنا	plot	منصوبہ بنانا
characterise	شناخت کرنا	practice	مشق کرنا
chase	پیچھا کرنا	preach	تلقین کرنا
cheer	حوصلہ افزائی کرنا	predict	پیشینگوئی کرنا
cite	حوالہ دینا	price	قیمت لگانا
claim	مطالبہ کرنا	proclaim	اعلان کرنا
command	کمان کرنا	project	تخمینہ لگانا
compare	موازنہ کرنا	promise	وعدہ کرنا
concern	احاطہ کرنا	promote	تشہیر کرنا
condemn	مزمت کرنا	prompt	ردعمل ابھارنا
conduct	رہنمائی کرنا	pronounce	تلفظ کرنا
confirm	تصدیق کرنا	propose	تجویز دینا
confront	سامنا کرنا	protect	حفاظت کرنا
contrast	مقابلہ کرنا	pursue	تعاقب کرنا
copy	نقل کرنا	quote	حوالہ دینا
correct	تصحیح کرنا	raise	پرورش کرنا
counter	سامنا کرنا	realise	احساس ہونا
cover	خبر دینا	reassure	ڈھارس بندھانا
decide	فیصلہ کرنا	rebuild	تعمیرنو کرنا
declare	اعلان کرنا	recommend	مشورہ دینا

defend	ساتھ دینا	register	اندراج کرنا
define	تعریف کرنا	regret	افسوس ہونا
demand	مطالبہ کرنا	regulate	نگرانی کرنا
depict	تصویرکشی کرنا	render	ترجمہ کرنا
detect	پتا لگانا	renew	تجدید کروانا
determine	فیصلہ کرنا	repair	مرمت کرنا
develop	وضاحت کرنا	repay	صلہ دینا
diagnose	تشخیص کرنا	report	خبر دینا
direct	سربراہی کرنا	represent	نمائندگی کرنا
discourage	حوصلہ شکنی کرنا	request	درخواست کرنا
display	نمائش کرنا	require	ضرورت ہونا
distinguish	شناخت بنانا	respect	عزت کرنا
divert	رخ موڑنا	restore	مرمت کرنا
donate	عطیہ دینا	review	جائزہ لینا
encounter	سامنا کرنا	revise	تصحیح کرنا
encourage	حوصلہ بڑھانا	reward	صلہ دینا
endorse	توثیق کرنا	ride	سواری کرنا
entertain	خاطرمدارت کرنا	round	چکر کاٹنا
envisage	تصورکرنا	sack	استیصال کرنا
estimate	اندازہ لگانا	sail	بحری سفر کرنا
exchange	تبادلہ کرنا	screen	طبی معائنہ کرنا
execute	قتل کرنا	seat	گنجائش رکھنا
exhibit	نمائش کرنا	secure	حفاظت کرنا

expect	توقع کرنا	sense	پتہ لگانا
exploit	استحصال کرنا	serve	خدمت کرنا
explore	جائزہ لینا	service	سروس کرنا
express	اظہارکرنا	spell	ہجہ کرنا
fight	مخالفت کرنا	split	کھال اتارنا
figure	حساب لگانا	sponsor	سرپرستی کرنا
fix	مرمت کرنا	stage	انتظام کرنا
follow	تعاقب کرنا	steer	قیادت کرنا
found	بنیاد رکھنا	stuff	پوست انبازی کرنا
found	بنیاد رکھنا	substitute	متبادل بننا
greet	استقبال کرنا	suggest	مشورہ دینا
guarantee	ضمانت دینا	summarize	خلاصہ کرنا
guard	رکھوالی کرنا	supervise	نگرانی کرنا
guess	اندازہ لگانا	support	حمایت کرنا
guide	رہنمائی کرنا	suppose	امید کرنا
head	سربراہی کرنا	survey	سروے کرنا
honour	عزت کرنا	suspect	شبہ کرنا
host	میزبانی کرنا	sustain	ہمت بندھانا
hunt	شکارکرنا	talk	بات کرنا
identify	شناخت کروانا	tour	دورہ کرنا
imagine	تصورکرنا	trace	سراغ لگانا
inherit	سامنا ہونا	trade	تجارت کرنا
initiate	آغازکرنا	translate	ترجمہ کرنا

inspect	معائنہ کرنا	transmit	ترسیل کرنا
inspire	احساس ابھارنا	treat	علاج کرنا
insure	بیمہ کرانا	try	کوشش کرنا
interpret	ترجمہ کرنا	undertake	ضمانت دینا
interrupt	تسلسل توڑنا	urge	تقاضا کرنا
interview	انٹرویو کرنا	value	قیمت لگانا
invest	سرمایہ کاری کرنا	visit	سیر کرنا
investigate	تفتیش کرنا	voice	اظہار کرنا
judge	اندازہ لگانا	watch	خیال رکھنا
justify	وضاحت کرنا	welcome	استقبال کرنا
launch	آغاز کرنا	witness	گواہی دینا
lead	رہنمائی کرنا		

Appendix A.5: List of verbs for R-5

Verb	Urdu Translation	Verb	Urdu Translation
accuse	پر الزام لگانا	govern	پر حکومت کرنا
amend	میں ترمیم کرنا	indicate	پر اشارہ کرنا
attack	پر حملہ کرنا	invoke	سے دعا مانگنا
believe	پر اعتبار کرنا	leap	سے چھلانگ لگانا
blame	پر الزام لگانا	line	میں استر لگانا
bless	پر رحمت فرمانا	love	سے محبت ہونا
capture	پر قبضہ کرنا	marry	سے شادی کرنا
comment	پر تبصرہ کرنا	modify	میں تبدیلی کرنا

consult	سے مشورہ کرنا	monitor	پر نظر رکھنا
contact	سے رابطہ کرنا	overcome	پر قابو پانا
counter	سے بچاؤ کرنا	process	پر کاروائی کرنا
delay	میں دیر کرنا	resign	سے استعفی دینا
discuss	پر بات کرنا	review	پر تبصرہ لکھنا
doubt	پر شک کرنا	sign	پر دستخط کرنا
ease	سے بوجھ اتارنا	slam	نکتہ چینی کرنا
emphasize	پر زور دینا	stress	پر زور دینا
exploit	سے فائدہ اٹھانا	suspect	پر شک کرنا
fish	میں مچھلی پکڑنا	talk	میں بات کرنا
flood	میں سیلاب لانا	voice	سے سر ملانا

Appendix A.6: List of verbs for verbs having XCOMP

Verb	**Urdu Translation**	**Verb**	**Urdu Translation**
advise	مشورہ دینا	permit	اجازت دینا
aim	ارادہ رکھنا	plan	منصوبہ بنانا
arrange	انتظام کرنا	pledge	عہد کرنا
attempt	کوشش کرنا	plot	سازش کرنا
bid	حکم دینا	pretend	دکھاوا کرنا
bother	زحمت کرنا	promise	عہد کرنا
choose	فیصلہ کرنا	request	درخواست کرنا
claim	دعویٰ کرنا	require	حکم دینا

command	حکم دینا	resolve	فیصلہ کرنا
decide	فیصلہ کرنا	seek	کوشش کرنا
direct	حکم دینا	signal	اشارہ کرنا
entitle	حق دینا	swear	قسم کھانا
expect	توقع کرنا	tempt	ترغیب دینا
free	وقت دینا	train	تربیت دینا
guarantee	ضمانت دینا	trouble	کوشش کرنا
influence	تحریک دینا	try	کوشش کرنا
instruct	حکم دینا	undertake	عہد کرنا
invite	دعوت دینا	urge	تقاضا کرنا
offer	پیشکش کرنا	venture	جرأت کرنا
order	حکم دینا		