Abstract

The thesis is an attempt to tackle the complex problem of Natural Language parsing using Lexical Functional Grammar (LFG). The language in question is Hindi which is a widely spoken and understood language in north and central India. Therefore, the utilitarian raison d'etre of the study will be suitably justified.

As part of the current research, grammar rules for Hindi using Lexical Functional Grammar (LFG) have been developed. In this formalism, the constraints are represented in the form of functional equations (f-equations). These equations augment the base component. The base component generates constituent structure (c-structure) which represents the word order in a sentence. The f-equations work as filters on sentences and generate a structure called functional structure (f-structure). An f-structure is a function mapping attributes to values. The attributes are grammatical relations and adjuncts. The values are atomic, predicates or the f-structure themselves.

The work takes off with the study of Hindi grammar. Hindi is a partially word order free language. Here, noun and adverb phrases occur in scrambled order in a sentence, but always precede the verb phrase. Within a phrase, however, the order of words is almost fixed.

We have followed the bottom-up approach for the design of the grammar. We studied the lexical and morphological properties of the words in the language. The words have been categorized, first into major syntactic categories, and subsequently depending upon the feedback, the words have been further subcategorized.

This categorization/subcategorization has been used to define and associate f-equations with lexical units and morphological processes.

Next, the base rules for different types of grammatical phrases were developed keeping in view the word order restrictions imposed by the language. The intra-phrase
agreement rules of the language have been realized by associating f-equations with the 
base rules. These f-equations along with the f-equations associated with lexical units 
work as filters for these phrases.

In the present work, we have come out with 17 rules for noun phrase, 118 rules for 
modifier phrase, 58 rules for verb phrase and 67 rules for adverb phrase.

Our analysis of simple Hindi sentences has yielded a corpus of 27 rules which 
accounts for these sentences. These rules have been derived in terms of phrasal rules. 
The rules for complex and compound sentences are obtained from the rules for simple 
sentences. The sentence level rules are augmented with f-equations to handle inter-
phrasal agreement properties imposed by the language.

The number of rules obtained for complex and compound sentences are 9 and 3 
respectively.

In the present work, special equations have been designed to handle immediate 
constraints and recursive generation of phrases. These are specific to Hindi and are 
not mentioned by Bresnan.

A rule-based parser has been developed by combining Tomita's parsing algorithm and 
LFG formalism. The output of this parser is an f-structure and a c-structure. The 
parser uses bottom-up parsing technique and generates c-structure and f-structure 
simultaneously. The system has been designed using Object Oriented approach and 
is implemented in C++. The system consists of a parsing table generation module, a 
parsing module and an f-structure generation module. This process of parsing has been 
supported by a lexicon.