

Transfer grammar rules : the case of relative clause construction in English, Hindi and Tamil

K. Lakshmi Bai

Language Technologies Research Centre
International Institute of Information Technology
(formerly Indian Institute Of Information Technology)
Hyderabad
(lakshmi@mail.iiit.net)

Abstract

In machine translation, a set of Transfer Grammar Rules are required for mapping syntactic representations of a source language, e.g. English, on to the target language representations, e.g. Hindi, Tamil etc. Such rules are required for simple and complex structures like relative clause construction. English has two major types of relative clauses, restrictive and non-restrictive, each with further structural variations. Indian languages such as Hindi and Tamil also have a number of such parallel constructions, some of which are common to them, e.g. participial and correlative relative clauses, and others language specific, e.g. vAA construction in Hindi.

The main concern of the paper is to examine the processes of relativization available in Hindi and Tamil from the perspectives of (i) stylistic preferences, (ii) structural constraints, and (iii) relative complexity of rule formulation, with a view to checking the feasibility of finding common transfer rules in machine translation from English to Hindi and Tamil. The data for the research will be drawn from descriptive accounts of Hindi and Tamil, experimentally elicited translation equivalents and machine accessible parallel corpora.

"Indian Languages and the Future of Information Technology"

Durgesh D Rao

Knowledge Based Computer Systems
National Centre for Software Technology
(durgesh@ncst.ernet.in)

Abstract

Human languages have so far played a peripheral role at best in mainstream information technology. And Indian languages have been even less in the picture. However, speech and language technology will have a central role to play in the information society of tomorrow, with its focus on speech interfaces and ubiquitous computing. And Indian languages can occupy a prominent place. This is not only due to their large population and potential market size. There is a lot that Indian languages can contribute to the rest of the world, with their strong oral tradition and rich cultural associations. For this idea to become a reality, linguists and technologists will have to work together as never before. In this talk, we will look at some of the possibilities for Indian languages, with two motivating examples -- a multilingual search engine, and a common phonetic encoding for all languages of the world.

The Preparation of Typological Grammars: The Need of the Hour

Karumuri V. Subbarao

Department of Linguistics

University of Delhi

(kvs2811@yahoo.com)

Abstract

In this paper I wish to focus my attention not so much on similarities between the innumerable languages of the South Asian subcontinent but on similarities and differences between the languages of a single family and between the language families of the subcontinent with the aim of proposing the notion of a typological grammar of a language. The commonalities between the languages have been discussed in detail by many scholars starting from Suniti Kumar Chatterji as early as 1926 in his magnum opus *The Origin and Development of the Bengali Language* (ODBL) and subsequently, by Emeneau (1956) who enunciated the idea of 'India as a linguistic area' with several insights that have been put forward in ODBL, and by Kuiper (1974). Just as the movies, TV soaps and commercials are Aryo-centric or Dravidio-centric, the data and analyses that are usually in vogue are either Indo-Aryan or Dravidian based with a neglect of the Tibeto-Burman or Munda languages or the Mon-Khmer Khasi (though there a few exceptions in recent years such as Anvita Abbi's work on reduplication). Such neglect could be due to the lack of available data or the typical Aryo-centric or Dravidio-centric attitude where every thing that is 'Indian' is either from the Indo-Aryan languages or from the Dravidian languages.

The study of the South Asian languages, just as the study of any language or languages, provides us a greater understanding of the nature of language. There are more than a thousand languages spoken in the South-Asian subcontinent and as is well known, they belong to four genetically different language families: Indo-Aryan, Dravidian, Tibeto-Burman and Mon-Khmer and Munda. Rarely does one find such a conglomeration of languages in one 'area' and coupled with intense bi/multilingualism that has been in existence in the subcontinent for thousands of years, the study of the syntax of these languages becomes much more fascinating. Though languages look

different at the surface level, they share a common Universal Grammar (UG) and the parameters they differ in are limited in number.

In this paper we wish to demonstrate how the commonalities and differences amongst the languages of the four language families can be fruitfully 'exploited' in the preparation of typological grammars of Indian languages.

An Online Electronic Dictionary of Idioms (Muhawara) and Proverbs (Lokokti) and Teacher Controlled Expert System

Pramod Kumar
Kendriya Hindi Sansthan
Delhi Centre
(pramod_khsd@sify.com)

Abstract

Online learning teaching refers to providing and accessing learning material through a computer via a network. The learning material is always available on network and is provided to a learner through learner's computer in response to his request. Keeping this in view along with the fact that it is a very difficult task to make a computer understand a natural language, but it can be used as a tool for teaching natural language, which at a higher level becomes learning of literature. To understand a literary piece one has to be aware of some basic concepts, and it includes learning of Muhawara (idiom) and lokokti (proverbs). Keeping this in view a paper design of online Artificially intelligent TCES is discussed in this paper. Once this bare system is developed and a dictionary of muhawara and lokokti is attached to it, it will help in teaching meaning and use of muhawara and lokokti. The bare system can be attached with other type of dictionaries to develop TCES of many other topics, like Word meaning, Antonyms, Synonyms etc. As this bare package is language independent it can be used to develop TCES of any language if the dictionaries are adopted to suit its needs.