8.20 Handling of Case Markers for Designing UNL Based Punjabi Language Server

Parteek Bhatia, Thapar Institute of Engineering and Technology, Patiala, Punjab

Abstract— Case marker plays an important role in designing a Language Server. In UNL based Machine Translation system Language servers are used to convert the source language text to UNL with the help of Enconverter and Deconverter is used to convert the UNL text to target language. In order to implement the complete multilingual machine translation on Internet, it requires the language server for each language which encompasses Enconverter and Deconverter for that language. In this paper we discuss the handling of case markers for designing Punjabi Language Server. It can play an important role in designing and developing Enconverter and Deconverter for Punjabi Language.

I. INTRODUCTION

The World Wide Web represents a formidable tool for communication and information access. With simple equipment, it is possible to access innumerable documents about a huge variety of topics, from any place around the world. However, despite the abundance of information, languages very often cause problems. [1] Most of the web pages today are written in few most commonly used languages like English, French, Chinese, etc.; it becomes difficult for a person with insufficient knowledge of these languages to access and use this tool of communication and information. This has prompted the need to devise means of automatically converting the information from one natural language to another natural language, called Machine Translation. [3] This process needs syntactic and semantic analysis of both source and target languages.

In case of Deconverter there are different phases for the generation of meaningful Punjabi sentences from the UNL. The process of deconversion involves syntax planning, case marker generation, and morphology phase. [2] After the syntax-planning phase, which is aimed at generation of proper sequence of words, case marking phase initiates. Case Marker phase used to express the complete contents of the sentence. In the deconverter process the first task is to parse the UNL file. The nodes are generated in the target language with the help of L-UW dictionary. After generation of target language nodes they are ordered in Syntax planning phase according to the grammatical details of target language. Then Case marker phase fills the nodes generated from syntax planning phase with the appropriate case marker for the target language depending upon the use of the relations in the UNL text. For this purpose it takes into consideration Relational Morphology of the target language.

II. UNIVERSAL NETWORKING LANGUAGE

Universal Networking Language (UNL), developed at UNU, is a formal language for representing the meaning of natural language sentences. The motivation behind UNL is to develop an interlingual representation such that semantically equivalent sentences of all languages have the same interlingual representation. Information expressed in UNL can be converted into the native user’s native language with higher quality and fewer mistakes than the computer translation systems. In addition UNL unlike natural language is free from ambiguities.

The UNL represents information sentence by sentence. Each sentence is converted into a directed hyper graph having concepts as nodes and relations as arcs.

The knowledge within document is expressed in three dimensions: [4]

- Word knowledge is expressed by Universal Words (UWs).
- Relating UWs through a set of UNL relations capture concept Knowledge.
- Speakers view, aspect, time of event, etc. are captured by UNL attributes.

Example UNL

Original sentence - John breaks the rules

UNL Corresponding to sentence is as follows:

```
agt(break(icr)-do).@entry,"John")
obj(break(icr)-do).@entry,rule@generic@pl)
```

In this, 'agt' and 'obj' are the relations where words starting with character "@" are attributes corresponding to each universal word, i.e. 'Break' is the universal word in this sentence and words like @pl, @generic are the attributes describing the universal word.

III. CASE MARKERS

Case is a category of morpho-syntactic properties which distinguishes the various relations that a noun phrase may bear to a governing head. Some of these relations are purely syntactic in nature. In the Indian linguistic system - descended from Sanskrit - the case constructs are called kaarakas. [5] As in the traditional understanding, they denote the relationship of the nominal with the main verb of the clause. The case structure in Punjabi is complex. An exhaustive study of the kaarak system with view to analyzing Punjabi into UNL has been carried out.

Case marker phase apply proper case marker for each and every relation in the given UNL expression.
There are total forty-five relations defined in UNL specifications and for each relation different case markers are used depending upon the grammatical details of that language.

Here we discuss the all forty-five relations defined in UNL [4] specifications with corresponding case markers for Punjabi language, which are used to design Case marker data file for Punjabi Language Server. Case marker for each relation in UNL is shown in table 1.

A Case Marker data file contains one or more set of constraints for each relation and each of these sets map to different case markers. So given a node with all its attributes including lexical attributes from dictionary, we search the database for appropriate rule which the node satisfies and accordingly the case markers are initialized for the case markers.

**Table 1 : Case Markers for UNL relations**

<table>
<thead>
<tr>
<th>No</th>
<th>UNL Relation</th>
<th>Description</th>
<th>Punjabi Case Marker</th>
<th>Example</th>
</tr>
</thead>
</table>
| 9  | cont | condition i.e., indicates a non-focused event or state that conditions a focused event or state | ने | ने कुछ बात कहते हैं तो उन्हें खुश बनाने के 
| 10 | cao | affected co-thing i.e., indicates a co-occurrent event or state for a focused event or state | छेड़, छोड़ | नाले के बीच खूब में किसी अन्य की भर्ति झुकाए जाते हैं 
| 11 | dur | duration i.e., indicates a period of time during which an event occurs or a state exists | रहेंगे | मीठे मीठे खाने का समय आ गया है 
| 12 | equ | affected co-thing i.e., indicates an equivalent concept | null | कभी भी रखी तो जीती है 
| 13 | frm | range/from-to i.e., indicates a range between two things | रहेंगे | परिसर तक उम्मीद करेंगे 
| 14 | frm | origin i.e., indicates an initial state of a thing or a thing initially associated with the focused thing | रहेंगे | परिसर तक उम्मीद करेंगे जिसका मूल है 
| 15 | gol | goal state i.e., indicates a final state of an object or a thing finally associated with the object of an event | हिस्ता | लांग परिसर विशेष उद्देश्य है 
| 16 | icl | indicates an upper concept or a more general concept | उर्दू संस्कृत हिन्दी बोलने वाला समझ रहा है 
| 17 | ins | Instrument i.e., indicates an instrument to carry out an event | टूट टूट स्पर्श टूट टूट हिटिये |
| 18 | int | Intersection i.e. indicates all common instances to have with a partner concept | null | जोड़ विवरण |
| 19 | iof | an instance of i.e. indicates a class concept that an instance belongs to | null | पाठिशाला का प्रतीक है |
| 20 | man | manner i.e. indicates a way to carry out an event or the characteristic of a state | null | नवरत्नी घटना |
| 21 | met | method or means i.e. indicates a way to carry out an event | ते घटना घटने के लिए विधि जिसना | पाठिशाल प्रतीक है |
| 22 | mod | modification i.e. indicates a thing that restricts a focused thing | कर चुका प्रकट करने का संशोधित वे | पाठिशाला उद नवरत्न |
| 23 | nam | name i.e. indicates a name of a thing | null | नवरत्नी वर्णन |
| 24 | obj | affected thing i.e. indicates a thing in focus that is directly affected by an event or state | ते घटना है प्रकट है जिसना घटना जवाबित करता | पाठिशाला प्रतीक है |
| 25 | opf | affected place i.e. indicates a place in focus affected by an event | ते निवासित वे परिसर | पाठिशाला प्रतीक है |
| 26 | or | disjunction i.e. indicates a partner to have disjunctive relation to | संयुक्त न हो | पाठिशाला प्रतीक है |
| 27 | per | proportion/rate distribution i.e. indicates a basis or unit of proportion, rate or distribution | देखिए घटना पुनः विश्वास | पाठिशाला प्रतीक है |
| 28 | plc | indicates a place where an event occurs, or a state that is true, or a thing that exists | ते घटना प्रकट है जिसना प्रतीक है |
| 29 | pif | initial place i.e. indicates a place where an event begins or a state that becomes true | ते पाठिशाला प्रतीक है जिसना प्रतीक है |
| 30 | plf | final place i.e. indicates a place where an event ends or a state that becomes false | ते पाठिशाला प्रतीक है जिसना प्रतीक है |
| 31 | pof | part of i.e. indicates a concept of which a focused thing is a part | ते पाठिशाला प्रतीक है जिसना प्रतीक है |
| 32 | pos | possessor i.e. indicates the possessor of a thing | null | पाठिशाला प्रतीक है |
| 33 | pth | partner i.e. indicates an indispensable non-focused agent of an action | ते पाठिशाला प्रतीक है जिसना प्रतीक है |
| 34 | pur | purpose i.e. indicates the purpose or objective of an event or the purpose of a thing that exists | ते पाठिशाला प्रतीक है जिसना प्रतीक है |
| 35 | qua | quantity i.e. indicates the quantity of a thing or unit | Null | पाठिशाला प्रतीक है |
| 36 | rsn | reason i.e. indicates a reason why an event or a state happens | ते पाठिशाला प्रतीक है जिसना प्रतीक है |
IV. RESULTS

In the following examples use of case markers are shown. From these examples we can understand how much case markers are important in natural language generation.

Example 1:
(Original sentence)

Ram ate dinner.

{unl}

agt (eat.@past.@entry, Ram (agt<person))

obj (eat.@post.@entry, dinner (icl<food))

{unl}

Output without case marker-

नम: भाज भाज

Now this sentence is not meaningful. To generate meaningful sentences we have to apply the case markers within the sentences. If "agt" has parent UW with @post attribute and it is transitive verb, and if child UW is noun, then put case marker ॐ after the child UW, i.e. Ram.

So ultimately the output after applying the case marker rule will be-

नम: भाज भाज

Example 2:
(Original sentence)

Post comes out from egg.

(Partial UNL)

src (come out(icl<happen)), @custom.@entry, egg(icl<foodstuff), @def)

Output without case marker rule-

वेंह आना आना आना ॐ

Again we can see that this sentence is meaning, so to generate a meaningful sentence case marker rule will be applied. It says that if child UW is noun and relation is 'src' (source) then put next case marker of child as ॐ.

Ultimately after applying the case marker rule output is:

वेंह आना ॐ आना आना ॐ
V. CONCLUSION

In this paper we discussed the case markers for the Punjabi language server based on UNL relations. In this we came to know how case markers are important for any language generation. We showed this with the help of examples. For any language, study of case constructs are very important for the generation of meaningful sentences.

REFERENCES


This paper was presented at LRIL-2007, National Seminar on Creation of Lexical Resources for Indian Language Computing and Processing at C-DAC Mumbai (26th to 28th March 2007), jointly organized by the Commission for Scientific and Technical Terminology (CSTT), New Delhi, MHRD, Govt. of India and the Centre for Development of Advanced Computing (C-DAC), Mumbai, Department of Information Technology, ME&IT, Govt. of India.