

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.

Parteek Bhatia is with Computer Science & Engineering Department, Thapar Institute of Engineering and Technology, Patiala, Punjab (e-mail: parteek.bhatia@tiet.ac.in)

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 interligual 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(icl>do).@entry,"John")
obj(break(icl>do).@entry, rule.@generic.@pl)
```

In this, 'agt' and 'obj' are the relations where words starting with character "@" are attributes corresponding to each universal words, 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 kaaraks. [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

No	UNL Relation	Description	Punjabi Case Marker	Example
1	agt	Agent i.e. a thing which initiates an action	ਨੇ	ਰਾਮ ਨੇ ਚੋਲ ਖਾਏ
2	and	conjunction i.e. a conjunctive relation between concepts	ਤੇ	ਰਾਮ ਤੇ ਸ਼ਾਮ ਦੇਸਤ ਹਨ
3	bas	basis i.e. indicates a thing used as the basis (standard) of comparison	ਤੇ	ਸੱਤ ਦਸ ਤੇ ਛੋਟਾ ਹੈ
4	ben	beneficiary i.e. indicates an indirectly related beneficiary or victim of an event or state	ਦੇ ਲਈ	ਦੇਸ਼ ਦੇ ਲਈ ਜਾਨ ਦੇਣਾ
5	cag	co-agent i.e. indicates a thing not in focus that initiates an implicit event that is done in parallel	ਦੇ ਨਾਲ	ਰਾਮ ਸ਼ਾਮ ਦੇ ਨਾਲ ਗਿਆ
6	cao	co-thing with attribute i.e. indicates a thing not in focus that is in a parallel state	ਦੇ ਨਾਲ	ਮਾਂ ਦੀਆਂ ਦੁਆਵਾਂ ਦੇ ਨਾਲ ਰਾਮ ਸੁਰਖਿਅਤ ਹੈ
7	ent	content i.e. indicates the content of a concept	null	ਪੈਸੇ ਬੰਦੇ ਜਾਣ ਦਾ ਖਤਰਾ
8	cob	affected co-thing	ਦੇ ਨਾਲ	ਰਾਮ ਮਿੱਠਰਾਂ ਦੇ ਨਾਲ ਦਿੱਲੀ ਗਿਆ

9	con	condition i.e. indicates a non-focused event or state that conditions a focused event or state	ਜੇ	ਜੇ ਤੁਸੀਂ ਬੱਕ ਗਏ ਹੋ ਤਾਂ ਅਸੀਂ ਘਰ ਚੱਲਦੇ ਹਾਂ
10	coo	effected co-thing i.e. indicates a co-occurrent event or state for a focused event or state	ਉਦੇ,ਜਦੋ	ਜਦੋ ਮੈਂ ਵਾਪਿਸ ਆਵਾਂ ਉਦੇ ਤੁਸੀਂ ਜਾਣਾ ਹੈ
11	dur	duration i.e. indicates a period of time during which an event occurs or a state exists	ਦੇ ਵੇਲੇ	ਸਵੇਰ ਦੇ ਵੇਲੇ ਮੈਸਮ ਚੰਗਾ ਹੁੰਦਾ ਹੈ
12	equ	effected co-thing i.e. indicates an equivalent concept	null	ਕਲਮ ਏਕ ਲਿਖਨੇ ਕੀ ਚੀਜ਼
13	fmr	range/from-to i.e. indicates a range between two things	ਤੋਂ	ਪਟਿਆਲਾ ਤੋਂ ਜਪਾਨ ਤਕ
14	frm	origin i.e. indicates an initial state of a thing or a thing initially associated with the focused thing	ਤੋਂ	ਪਟਿਆਲਾ ਤੋਂ ਦਿੱਲੀ ਦੂਰ ਹੈ
15	gol	goal state i.e. indicates a final state of object or a thing finally associated with the object of an event	ਵਿਚ	ਰਾਮ ਪਟਿਆਲਾ ਵਿੱਚ ਰਹਿੰਦਾ ਹੈ
16	icl	indicates an upper concept or a more general concept	ਤਰਾਂ ਦਾ	ਕੁੱਤਾ ਏਕ ਤਰਾਂ ਦਾ ਜਾਨਵਰ ਹੈ
17	ins	Instrument i.e. indicates an instrument to carry out an event	ਦੇ ਨਾਲ	ਕਲਮ ਦੇ ਨਾਲ ਲਿਖੋ

18	int	Intersection i.e. indicates all common instances to have with a partner concept	null	ਇਕ ਕਿਤਾਬ
19	iof	an instance of i.e. indicates a class concept that an instance belongs to	null	ਪਟਿਆਲਾ ਪੰਜਾਬ ਦਾ ਸ਼ਹਿਰ ਹੈ
20	man	manner i.e. indicates a way to carry out an event or the characteristics of a state	null	ਜਲਦੀ ਚਲੋ
21	met	method or means i.e. indicates a means to carry out an event	ਦੇ ਨਾਲ	ਪੱਥਰ ਹਥੜੇ ਦੇ ਨਾਲ ਤੋੜਿਆ ਗਿਆ
22	mod	modification i.e. indicates a thing that restricts a focused thing	ਦਾ	ਚੰਦਰਮਾ ਧਰਤੀ ਦਾ ਉਪਗ੍ਰਹਿ ਹੈ
23	nam	name i.e. indicates a name of a thing	null	ਓਹ ਰਾਮ ਹੈ
24	obj	affected thing i.e. indicates a thing in focus that is directly affected by an event or state	ਤੋਂ	ਰਾਮ ਨੇ ਸ਼ੀਤਾ ਤੋਂ ਕੰਮ ਕਰਾਇਆ
25	opl	affected place i.e. indicates a place in focus affected by an event	ਤੇ	ਵਿਚਕਾਰ ਤੋਂ ਕੱਟੋ
26	or	disjunction i.e. indicates a partner to have disjunctive relation to	ਜਾਂ	ਤੁਸੀਂ ਅੱਜ ਇਥੇ ਰਹੋਗੇ ਜਾਂ ਨਹੀਂ
27	per	proportion/rate/distribution i.e. indicates a basis or unit of proportion, rate or distribution	ਵਾਰ	ਮੈਂ ਚਾਰ ਵਾਰ ਉਥੇ ਗਿਆ

28	plc	indicates a place where an event occurs, or a state that is true, or a thing that exists	ਵਿਚ	ਰਸੋਇ ਵਿੱਚ ਖਾਨਾ ਬਨ ਰਿਹਾ ਹੈ
29	plf	initial place i.e. indicates a place where an event begins or a state that becomes true	ਤੇ	ਪਟਿਆਲਾ ਤੇ ਜਪਾਨ ਤਕ
30	plt	final place i.e. indicates a place where an event ends or a state that becomes false	ਤਕ	ਪਟਿਆਲਾ ਤਕ ਸਫਰ
31	por	part of i.e. indicate a concept of which a focused thing is a part	ਦਾ/ਦੀ/ਦੇ	ਕਿਤਾਬ ਰਾਮ ਦੀ ਹੈ
32	pos	possessor i.e. indicates the possessor of a thing	null	ਮੇਰੀ ਕਿਤਾਬ
33	ptn	partner i.e. indicates an indispensable non-focused initiator of an action	ਦੇ ਨਾਲ	ਰਾਮ ਦੇ ਨਾਲ ਸ਼ਾਮ ਰਹਿੰਦਾ ਹੈ
34	pur	purpose i.e. indicates the purpose or objective of an agent of an event or the purpose of a thing that exists	ਦੇ ਲਈ	ਰਾਮ ਨੂੰ ਬਚਾਉਣ ਦੇ ਲਈ ਸ਼ਾਮ ਝੂਠ ਬੋਲ ਰਿਹਾ ਹੈ
35	qua	quantity i.e. indicates the quantity of a thing or unit	Null	ਦੇ ਚੂਹੇ
36	rsn	reason i.e. indicates a reason why an event or a state happens	ਦੇ ਕਰਨ	ਮੀਂਹ ਦੇ ਕਾਰਨ ਸਾਰੀ ਫਸਲ ਖਰਾਬ ਹੋ ਗਈ

37	scn	scene i.e. indicates a scene where an event occurs, or state is true, or a thing exists	ਵਿਚ	ਰਾਮ ਦੀ ਤਸਵੀਰ ਕੱਲ ਟੀ ਵੀ ਵਿੱਚ ਦਿਖਾਈ ਗਈ
38	seq	sequence i.e. Indicates a prior event or state of a focused event or state	ਦੇ ਬਾਅਦ	ਤੀਨ ਦੇ ਬਾਅਦ ਚਾਰ
39	src	indicates the initial state of an object or thing initially associated with the object of an event	ਤੋਂ	ਬਤੀ ਲਾਲ ਤੋਂ ਹਰੀ ਹੋਈ
40	tim	time i.e. indicates the time an event occurs or a state is true	ਨੂੰ	ਮੰਗਲਵਾਰ ਨੂੰ ਚਲਾਂਗੇ
41	tmf	initial time i.e indicates the time an event starts or a state becomes true	ਤੋਂ	ਸਵੇਰ ਤੋਂ ਲੈਕੇ ਸ਼ਾਮ ਤਕ
42	tml	final time i.e indicates a time an event ends or a state becomes false	ਤਕ	ਸ਼ਾਮ ਤਕ ਘਰ ਪਹੁੰਚਾਂਗੇ
43	to	indicates a final state of a thing or a final thing (destination) associated with the focused thing	ਦੇ ਲਈ	ਗਡੀ ਲੰਦਨ ਦੇ ਲਈ ਹੈ
44	via	an intermediate place or state	ਵਲੋਂ	ਪਟਿਆਲਾ ਵਲੋਂ ਜਲਦੀ ਪਹੁੰਚਾਂਗੇ
45	aoj	A thing which is in a state or has an attribute	Null	ਲਾਲ ਕਲਮ

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.@past.@entry, dinner (icl<food))
{/unl}

Output without case marker-

ਰਾਮ ਖਾਣਾ ਖਾਧਾ

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 *@past* attribute and it is transitive verb, and if child UW is noun, then put case marker ਨੇ after the child UW, i.e. Ram.

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

ਰਾਮ ਨੇ ਖਾਣਾ ਖਾਧਾ

Example 2:
(Original sentence)

Pest comes out from egg.

(Partial UNL)
src(come out(icl>happen). @custom.@entry, egg(icl>foodstuff).@def)

Output without case marker rule-

ਬਚਾ ਅੰਡੇ ਬਾਹਰ ਆਉਂਦਾ ਹੈ

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 ਤੋਂ .

Ultimately after applying the case marker rule output is:

ਬਚਾ ਅੰਡੇ ਤੋਂ ਬਾਹਰ ਆਉਂਦਾ ਹੈ

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

- [1] P. Bhattacharyya Multilingual Information Processing Using Universal Networking Language, Indo UK Workshop on Language Engineering for South Asian Languages (LESAL), Mumbai, India, April, 2001.
- [2] Ritesh Kumar Sinha: Hindi Generation : Syntax Planing and Case marking, Mini Project Report, 2005
- [3] S., Parikh J. and Bhattacharyya P: 2002, Interlingua Based English Hindi Machine Translation and Language Divergence, journal of Machine Translation (JMT), Volume 17
- [4] The Universal Networking Language (UNL): Specifications Version 3 Edition 2, <http://www.unl.org/unlsys/unl/UNL%20Specifications.htm>
- [5] Kuntal dey and pushpak Bhattacharya: UNL based analysis and generation of Bengali case structure constructs, <http://www.cfil.itb.ac.in/convergence03/all%20data/paper%20032-37.pdf>

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, MC&IT, Govt. of India.