

A Frame Work for Designing Punjabi Deconverter & Case Markers for Punjabi Language Server

By

Parteek Bhatia

Faculty, Department of Comp Sc & Engg
THAPAR University, Patiala (Punjab)



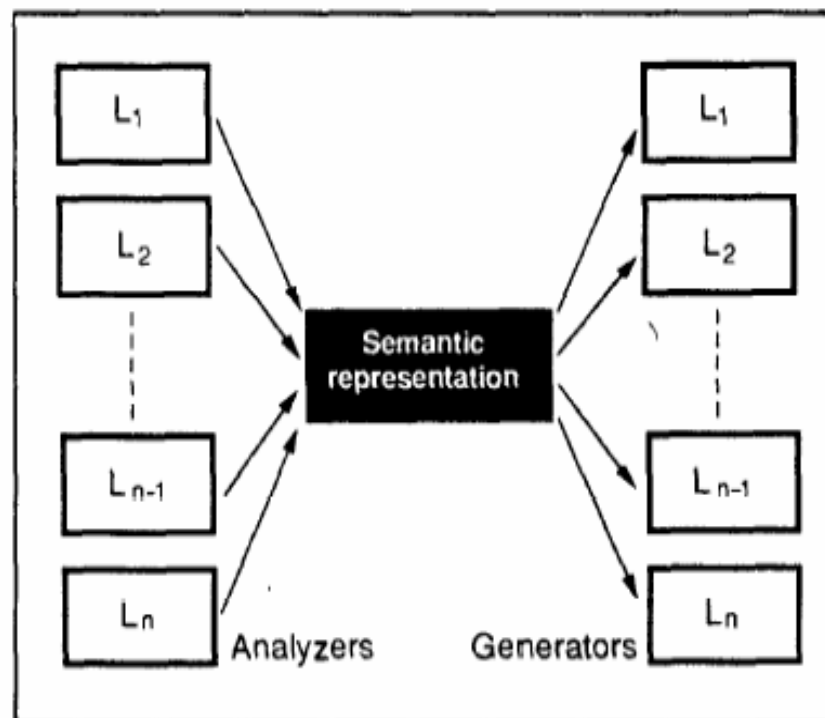
Machine Translation

The automated machine translation system is one of the dreams of human beings.

Machine Translation (MT) is used for computerized methods that automate all or part of the process of translating from one human language to another.

Interlingua Model for MT

An interlingua system is much easier to extend by adding new language pairs, than a transfer system. It requires two components for each language, one for conversion of native language to interlingua and other for interlingua to native language. Thus, it requires $2n$ components instead of $(n*(n-1))$ components in case of transfer based approach.



Use of UNL as Interlingua

In 1990, the Institute of Advanced Studies of the United Nations University, Tokyo began its multinational interlingua based MT project – based on a **‘standardized’ intermediary language, Universal Networking Language (UNL)**.

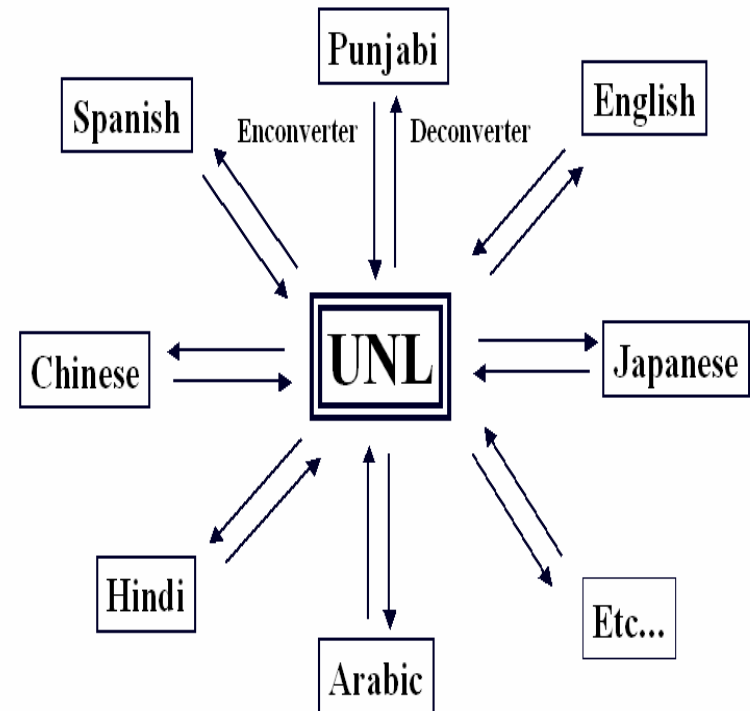
The UNL is an electronic language for computers to express and exchange every kind of information.

The motivation behind UNL is to develop an interlingua representation such that semantically equivalent sentences of all languages have the same interlingua representation.

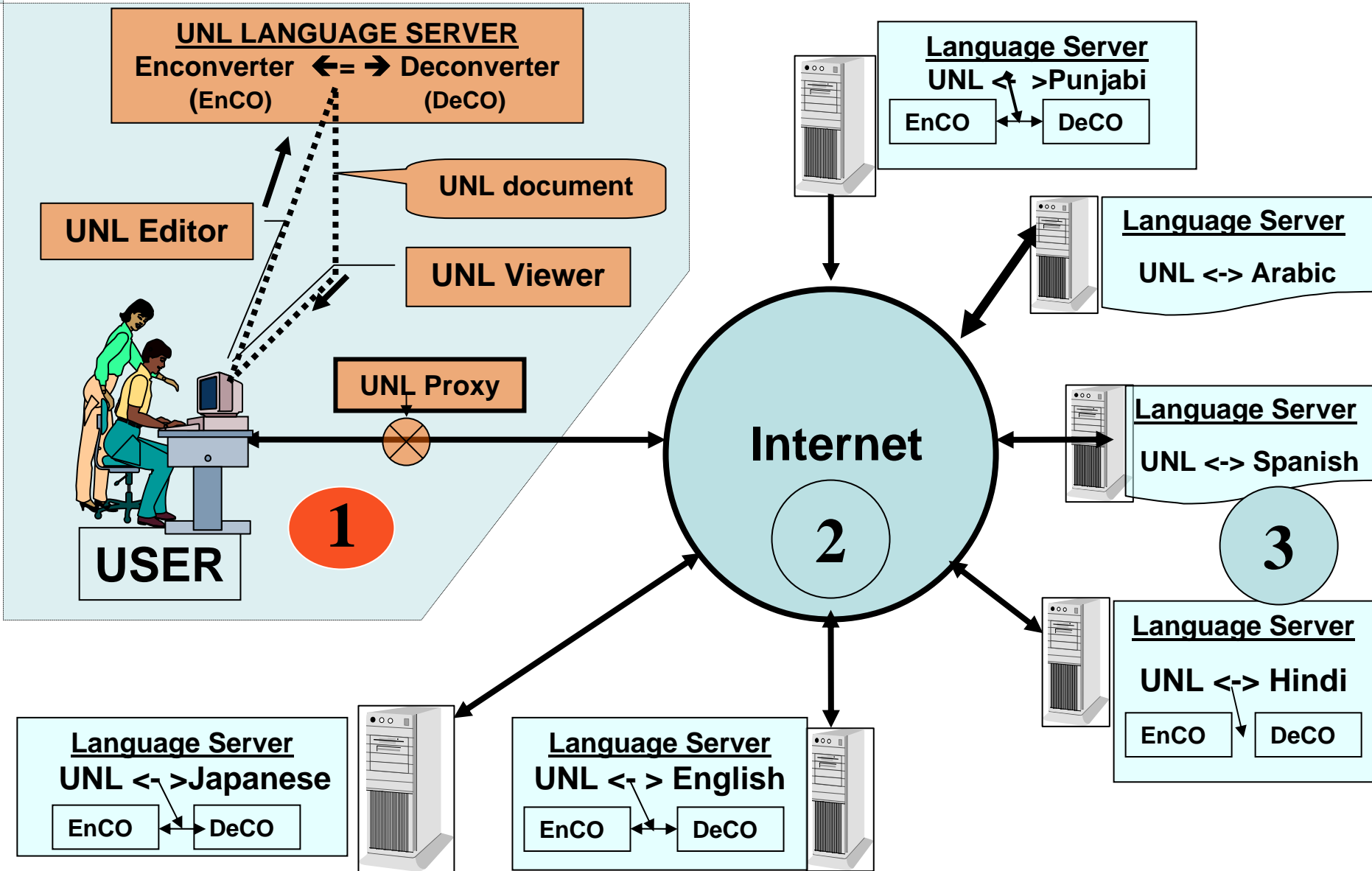
Language Server

There are two important modules of UNL based Language Server.

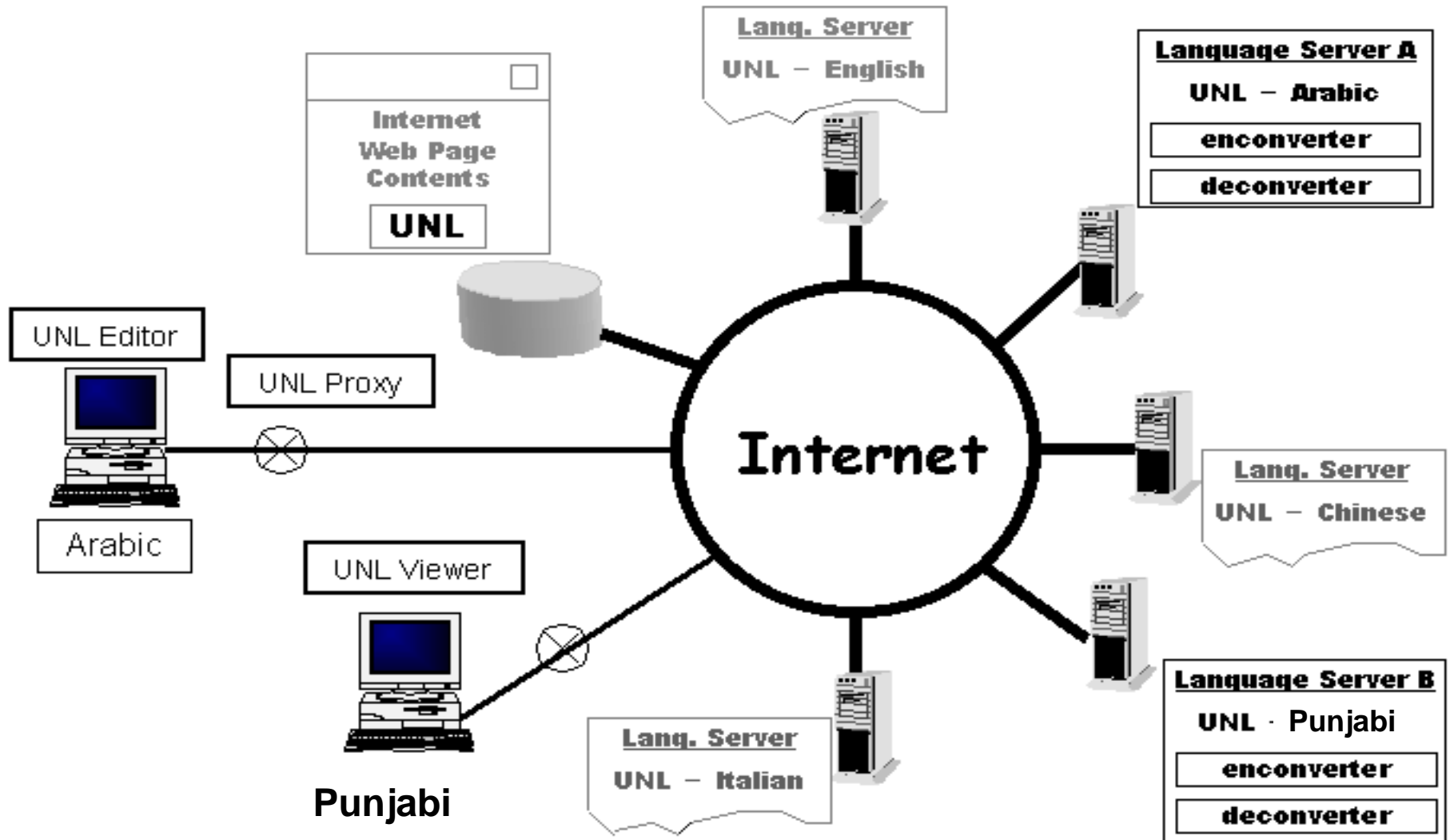
- **Enconverter**
It processes input text and produces the UNL as output.
- **Deconverter**
It takes UNL as input and produces text in native language.



The UNL Over Web



The UNL Over Web



UNL's Representation of Information

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:

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

UNL Constituents

Universal Word

Universal Words (UWs) are character-strings used to represent simple or compound concepts. It is made up of a character string followed by a list of constraints

<UW> ::= <Head Word> [<Constraint List>]

Head Word

Head Word is an English word / Compound Word / Phrase or Sentence that is interpreted as a label for a set of concepts.

For example drink, eat, dog *etc.* can act as Head Word.

Constraints or Restrictions

The Constraint List restricts the range of the concept that a basic UW represents.

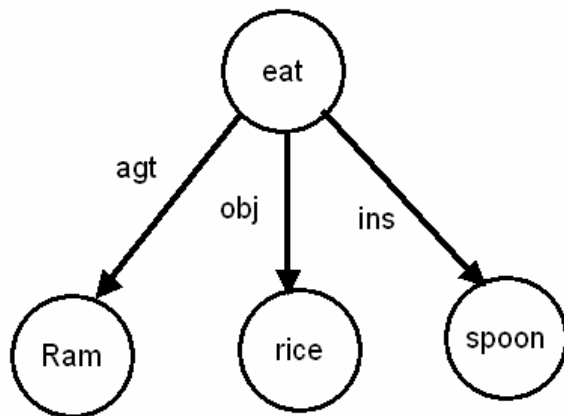
For example

[ਦੇ ਸ਼] ::= state(equ>nation)	:denotes nation,
[ਹਾ ਲਤ] ::= state(icl>situation)	: kind of situation,
[ਸਰਕਾ ਰ] ::= state(icl>government)	:kind of government

UNL Constituents Contd..

Relations

One UNL construct is made up of one relation and two UWs. Example: Ram is eating rice with a spoon.



Relation	Description
agt	agent: a thing which initiates an action
bas	basis: a thing used as the basis
con	condition: a non-focused event which conditioned a focused event
dur	duration: a period of time during an event occurs or a state exists
fnt	range: a range between two things
icl	a kind of: a more general concept
man	manner: the way to carry out event or characteristics of a state
nam	name: a name of a thing
obj	affected thing: a thing in focus which is directly affected by an event
opl	affected place: a place in focus where an event affects
qua	quantity: a quantity of a thing or unit
rsn	reason: a reason that an event or a state happens
tmf	initial time: the time an event starts or a state becomes true

UNL Constituents Contd..

Attributes

Attributes of UWs are used to describe subjectivity of sentences. They show what is said from the speaker's point of view.

Concept	Attributed As
Time with respect to the speaker	@past, @present, @future
Speaker's view of Aspect	@begin, @complete, @continue
Speaker's view of Reference	@generic, @def, @indef
Speaker's Focus	@emphasis, @entry, @qfocus
Speaker's attitudes	@confirmation, @exclamation, @interrogative
Speaker's view point	@ability, @although, @conclusion, @doubt
Describing speaker's attitudes	@polite, @request

UNL Constituents Contd..

Example:

Ram eats mango.

UNL's representation:

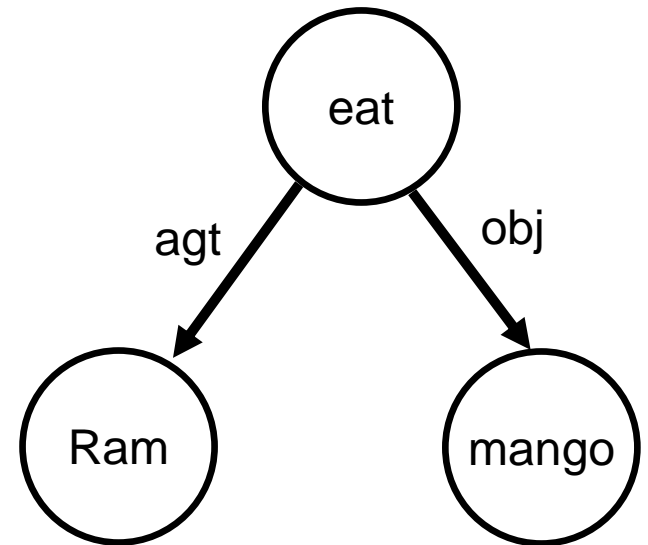
```
{unl}
```

```
agt(eat. @present. @entry, Ram)
```

```
obj(eat. @present. @entry, mango)
```

```
{/unl}
```

Here, agt indicates a thing which initiates an action and obj indicates affected thing.



Punjabi Deconverter Architecture

A Punjabi Deconverter will consist of word dictionary and conversion rules for a Punjabi language. This engine takes UNL expression as input and generates target language (Punjabi) sentence with the help of various database files like lexicon files, morphological rule files *etc.*

There are basically four modules for UNL Punjabi Deconversion:

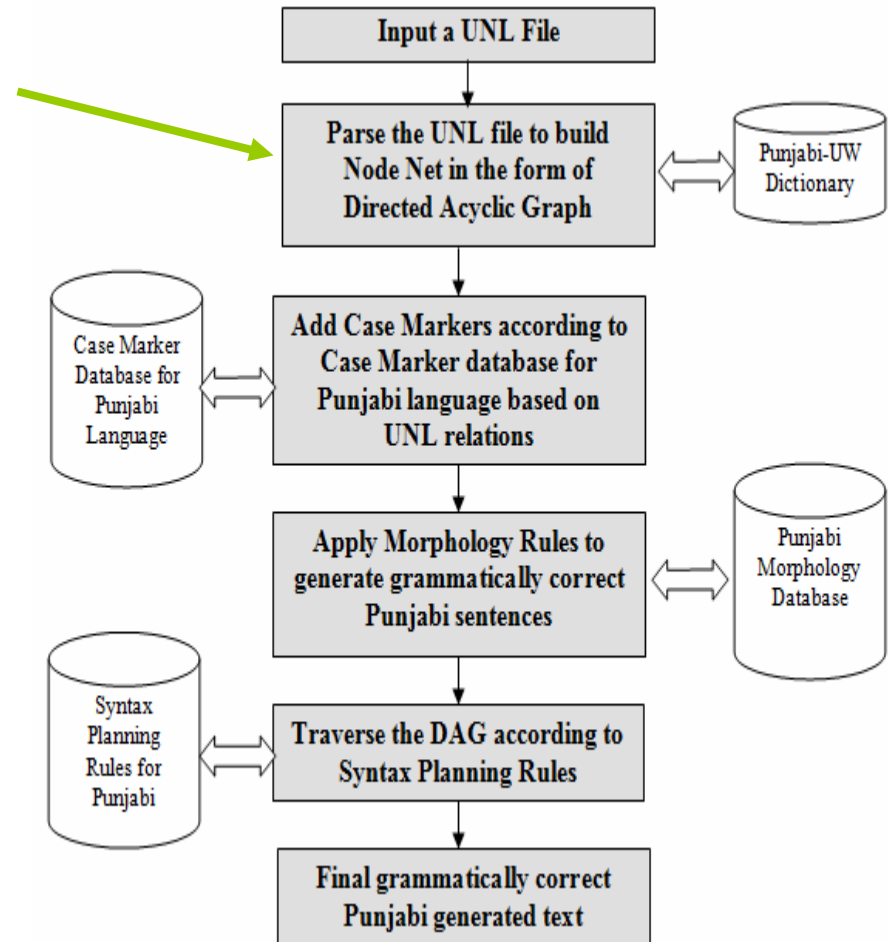
- UNL Parser
- Case-Marking Module
- Morphology Generation Module
- Syntax Planning Module

Punjabi Deconverter Architecture Contd..

UNL Parser

It is used to read an UNL file and convert it into machine understandable format and it performs the following tasks:

- It reports certain errors, if exists, in the input UNL file.
- It instantiates the nodes, scope-nodes and relations present in UNL document.
- It build the node-net for every sentence present in UNL document.



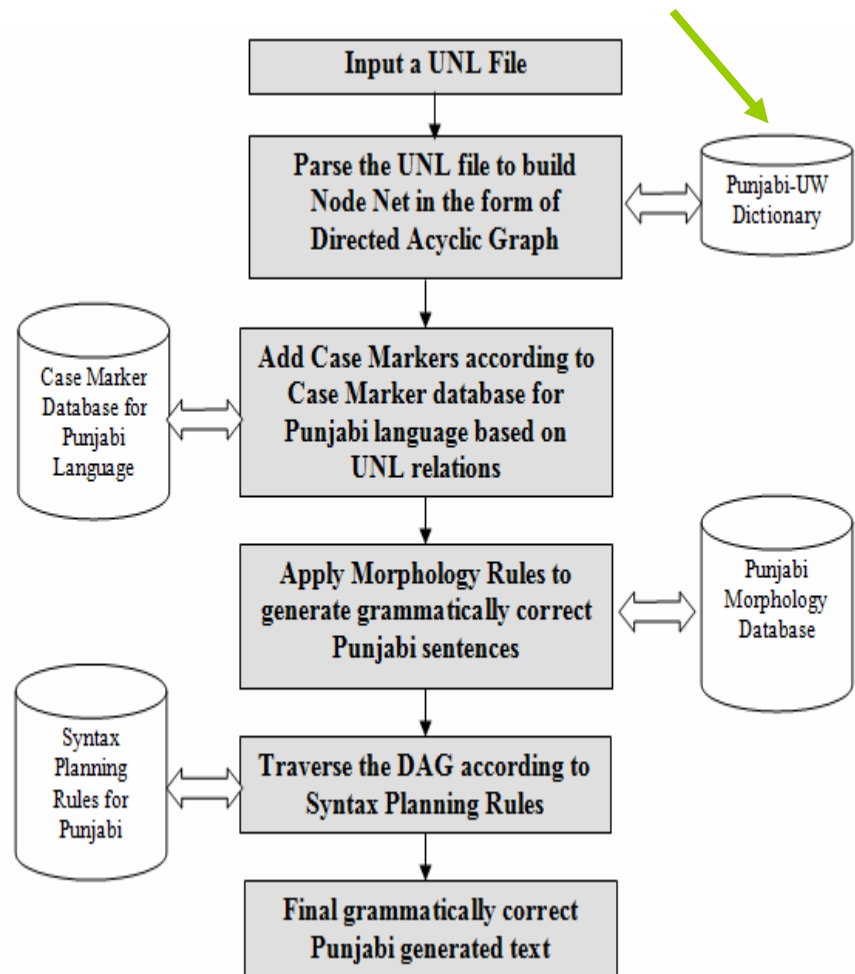
Punjabi Deconverter Architecture Contd..

Punjabi-UW Dictionary

It is used to replace the English head words in UNL file to equivalent Punjabi words.

<UW> ::= <Head Word>
[<Constraint List>]

[ਦਾ ਖ਼ਲਾ] ::= "admission(ict>action)"
(N,MALE,INANI,NA); Admission is prohibited.

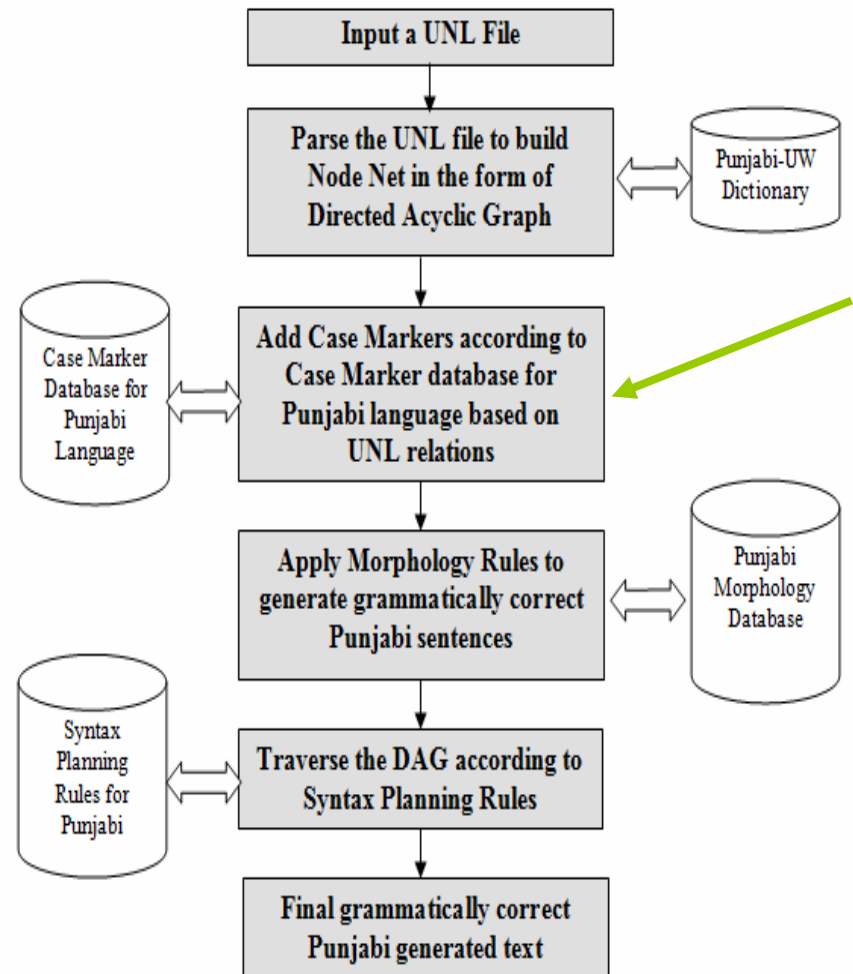


Punjabi Deconverter Architecture Contd..

Case Marking Phase

Case marker module apply proper case marker for each and every relation in the given UNL expression. We follow a rule base approach to incorporate the case markers correctly. It takes care of what Case Marker should precede or follow the individual words.

For example, words like ਨੇ , ਤੇ , ਦੇ ਲਈ, ਦੇ ਨਾਲ are added with Punjabi words.



Punjabi Deconverter Architecture Contd..

Morphology Phase

To get the more natural meaning of the word, the word should be changed or something should be removed or added to get the complete sense of the sentence. e.g.

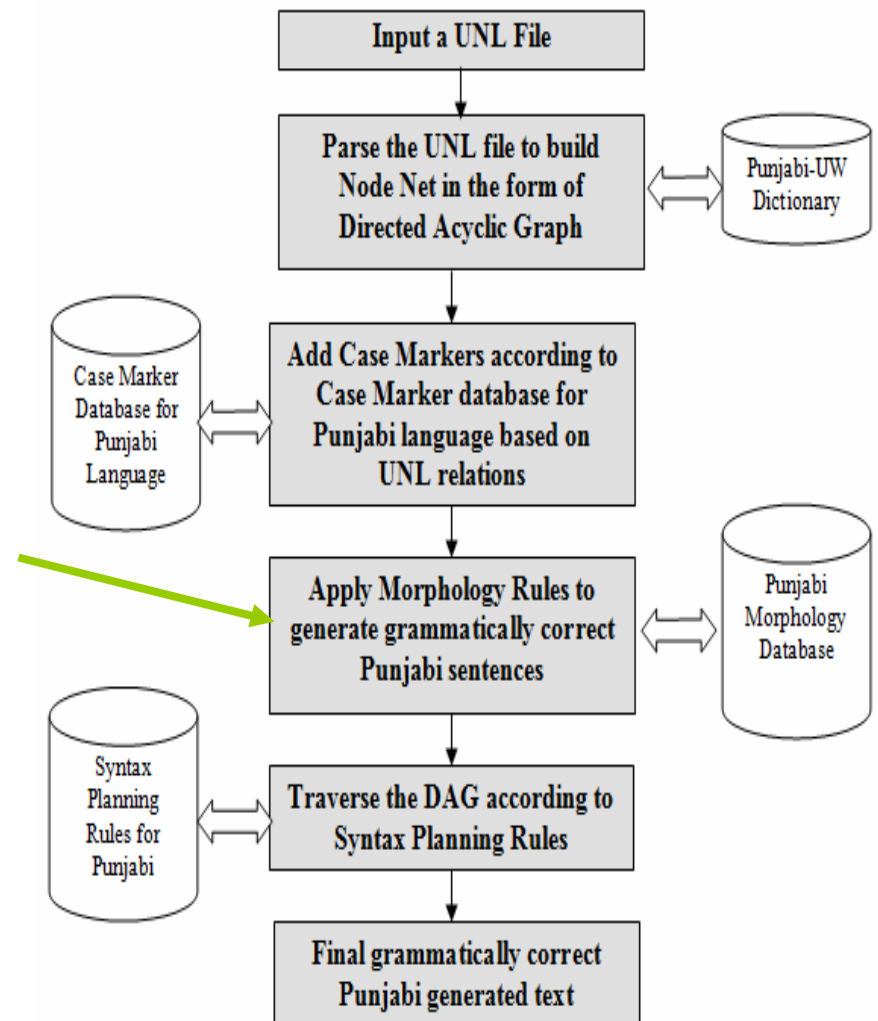
ਬੱਚਾ ਖਾਣਾ ਖਾ ਰਿਹਾ ਹੈ ।
ਬੱਚੀ ਖਾਣਾ ਖਾ ਰਹੀ ਹੈ ।

Here, ਬੱਚ becomes ਬੱਚਾ and in case of male and ਬੱਚੀ in case of female. Similarly ਰਹ becomes ਰਿਹਾ and ਰਹੀ for male and female respectively.

For example

If @progress, @present, @sg, @male in UNL

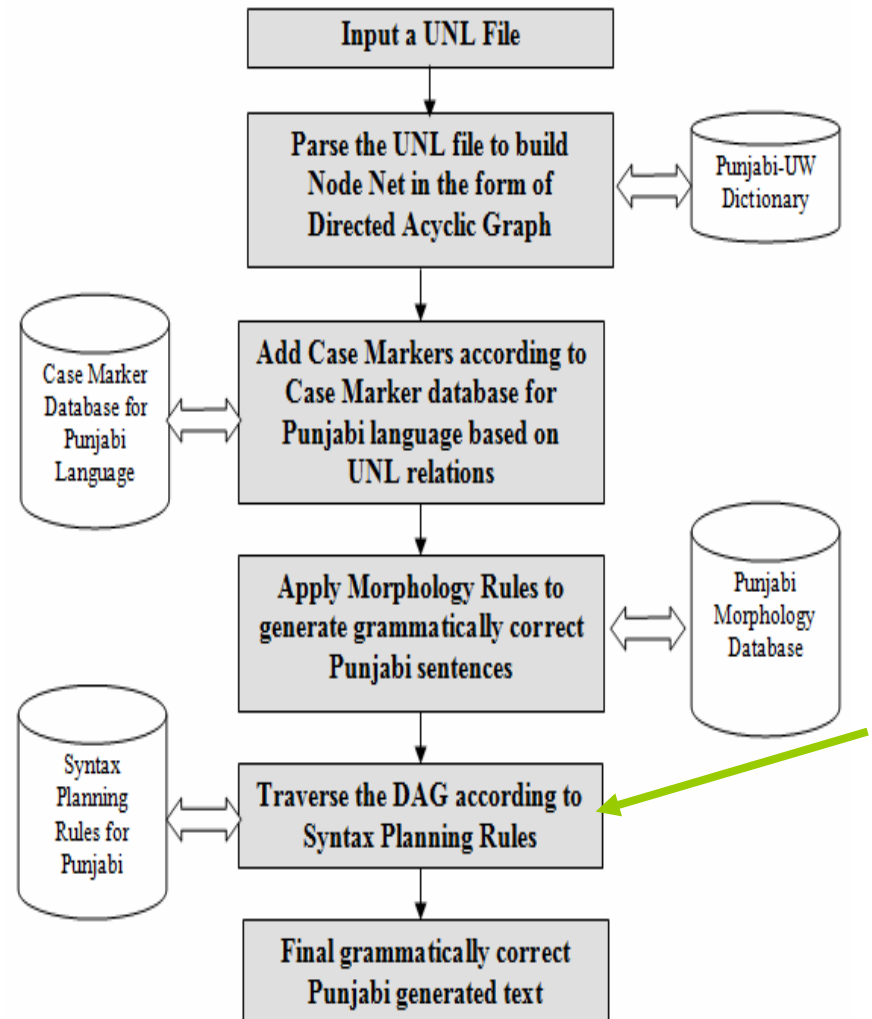
then ਰਿਹਾ ਹੈ is added with the verb



Punjabi Deconverter Architecture Contd..

Syntax Planning Phase

It generates proper sequence of words for the target sentence. It traverse the Directed Acyclic Graph (DAG) structure or Node net generated by UNL Parser. Proper traversal of the node net generates the syntax plan of the target sentence.



Matrix Based Syntax Planning

- A matrix with dimension 46×46 (for each UNL relation) is created. Let the Matrix be M . Then any element of the matrix will be denoted by M_{ij} where M_{ij} is element of the i th row and j th column. If the $M_{ij} = L$ then it means that the position of the child of the i th relation label is left of the child of the j th relation label. Similarly $M_{ij} = R$ then it means that the position of the child of the i th relation label is right of the child of the j th relation label.
- A rank for each relation label is calculated by adding the number of 'R' in the row of each relation label. The higher the value of the rank lesser is the priority.
- It helps to make language independent Deconverter.

Examples: Matrix Based Syntax Planning Contd..

English: Ram is eating rice with a spoon.

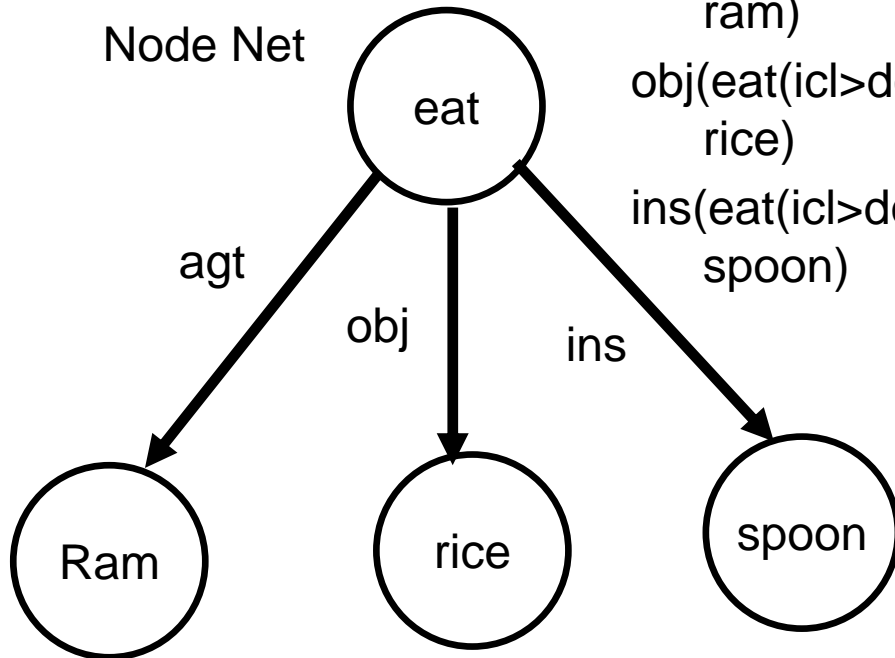
Punjabi: ਰਾ ਮ ਚਮਚ ਦੇ ਨਾ ਲ ਚੈ ਲ ਖਾ ਿ ਰਹਾ ਹੈ

UNL:

agt(eat(icl>do).@progress,@present,@sg,@male,@entry,ram)

obj(eat(icl>do).@progress,@present,@sg,@male,@entry,rice)

ins(eat(icl>do).@progress,@present,@sg,@male,@entry,spoon)



ਰਾ ਮ ਚਮਚ ਚੈ ਲ ਖਾ

Priority Matrix

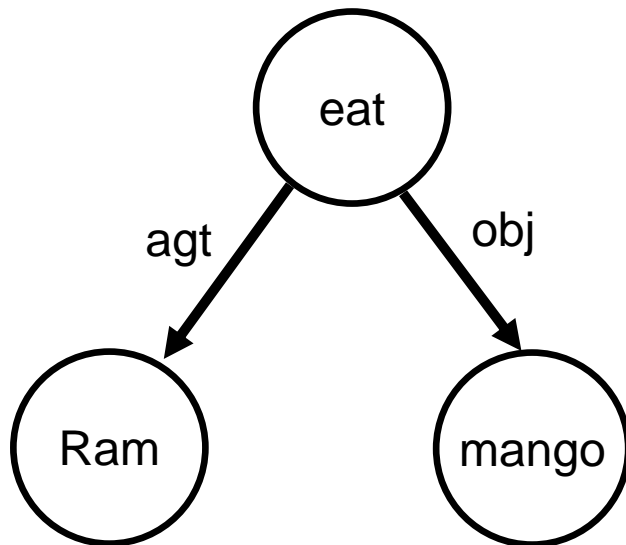
	agt	obj	ins
agt	-	L	L
obj	R	-	R
ins	R	L	-

Examples: Matrix Based Syntax Planning Contd..

English: Ram eats mango.

Punjabi: ਰਾਮ ਅੰਬ ਖਾਂਦਾ ਹੈ

UNL: agt(eat.@present.@entry,Ram)
obj(eat.@present.@entry,mango)



Node Net

	agt	obj
agt	-	L
obj	R	-

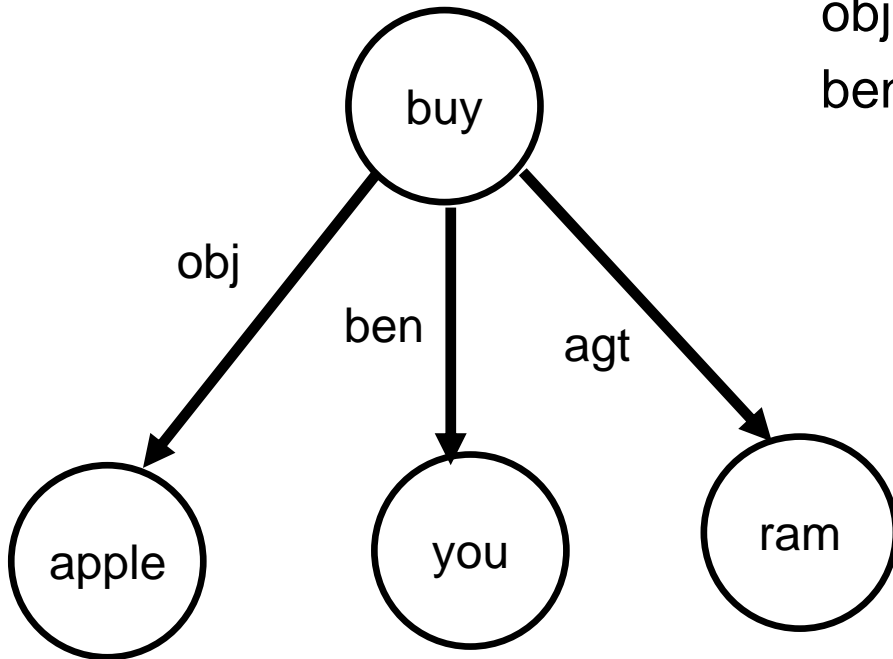
Priority Matrix

Examples: Matrix Based Syntax Planning Contd..

English: Ram bought an apple for you.

Punjabi: ਰਾ ਮਨੇ ਤੇ ਰੇ ਲਈ ਸੇ ਬਖਰੀ ਿ ਦਆ

UNL: agt(buy.@entry.@past, Ram)
 obj(buy.@entry.@past, apple.@def)
 ben(buy.@entry.@past, you)



Node Net

	agt	obj	ben
agt	-	L	L
obj	R	-	R
ben	R	L	-

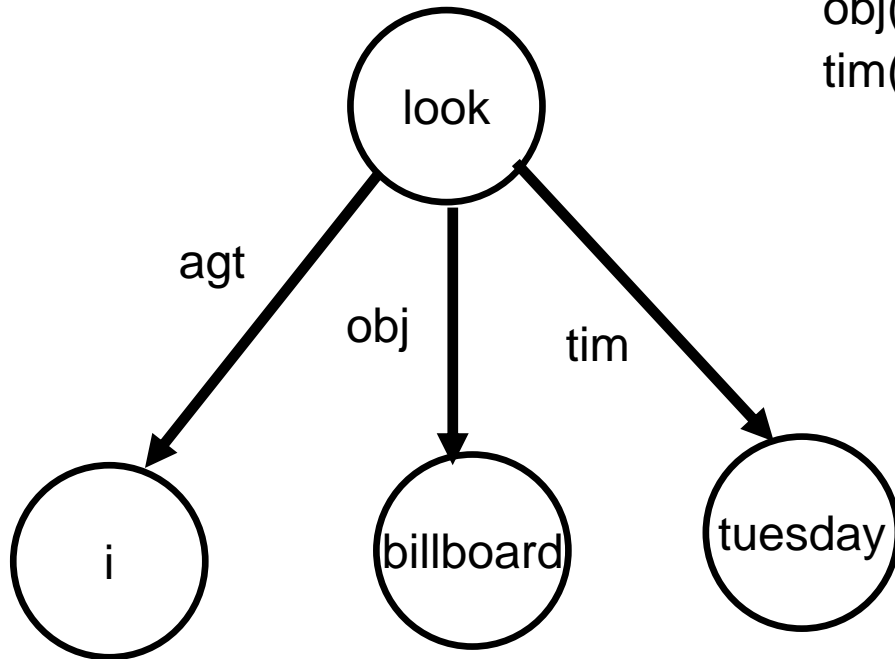
Priority Matrix

Examples: Matrix Based Syntax Planning Contd..

English: I look at the billboard on Tuesday.

Punjabi: ਮੈਂ ਮੰ ਗਲਵਾ ਰ ਨੂੰ ਇਸ਼ਾ ਤਹਾ ਰ ਵੱ ਲ
ਦੇ ਖਦਾ ਰਾ

UNL: `agt(look(icl>do).@entry, i)`
`obj(look(icl>do).@entry, billboard.@def)`
`tim(look(icl>do).@entry, tuesday(icl>time).@def)`



Node Net

	agt	obj	tim
agt	-	L	L
obj	R	-	R
tim	R	L	-

Priority Matrix

Role of Case Marker for Designing Language Server

- Case marker plays an important role in designing a Language Server. 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 marker are used depending upon the grammatical details of that language.

Punjabi Case Markers

Correspondence with UNL Relations

Sr No	UNL Relation	Description	Punjabi Case Marker	Example
1	agt	Agent i.e. a thing which initiates an action	ਨੇ	ਰਾਮ ਨੇ ਚੌਲ ਖਾਧੇ
2	and	conjunction i.e a conjunctive relation between concepts	ਤੇ	ਰਾਮ ਤੇ ਸ਼ਾਮ ਦੋਸਤ ਹਨ
3	aoj	A thing which is in a state or has an attribute	Null	ਲਾਲ ਕਲਮ
4	bas	basis i.e. indicates a thing used as the basis (standard) of comparison	ਤੋਂ	ਸੱਤ ਦਸ ਤੋਂ ਛੋਟਾ ਹੈ
5	ben	beneficiary i.e indicates an indirectly related beneficiary or victim of an event or state	ਦੇ ਲਈ	ਦੇਸ਼ ਦੇ ਲਈ ਜਾਨ ਦੇਣਾ
6	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

Punjabi Case Markers Correspondence with UNL Relations Contd..

7	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 ਸੁਰਖਿਅੱਤ ਹੈ
8	cnt	content i.e. indicates the content of a concept	null	ਪੈਸੇ ਖੋਏ ਜਾਣ ਦਾ ਖਤਰਾ
9	cob	affected co-thing	ਦੇ ਨਾਲ	ਰਾਮ ਮਿੱਤਰਾਂ ਦੇ ਨਾਲ ਦਿੱਲੀ ਗਿਆ
10	con	condition i.e. indicates a non-focused event or state that conditions a focused event or state	ਜੇ	ਜੇ ਤੁਸੀਂ ਥੱਕ ਗਏ ਹੋ ਤਾਂ ਅਸੀਂ ਘਰ ਚੱਲਦੇ ਹਾਂ
11	coo	effected co-thing I.e. indicates a co-occurrent event or state for a focused event or state	ਉਦੋ,ਜਦੋ	ਜਦੋ ਮੈਂ ਵਾਪਿਸ ਆਵਾਂ ਉਦੋ ਤੁਸੀਂ ਜਾਣਾ ਹੈ
12	dur	duration i.e indicates a period of time during which an event occurs or a state exists	ਦੇ ਵੇਲੇ	ਸਵੇਰ ਦੇ ਵੇਲੇ ਮੌਸਮ ਚੰਗਾ ਹੁੰਦਾ ਹੈ

Punjabi Case Markers Correspondence with UNL Relations Contd..

13	equ	effected co-thing i.e. indicates an equivalent concept	null	ਕਲਮ ਏਕ ਲਿਖਨੇ ਕੀ ਚੀਜ਼
14	fmt	range/from-to i.e. indicates a range between two things	ਤੋਂ	ਪਟਿਆਲਾ ਤੋਂ ਜਪਾਨ ਤਕ
15	frm	origin i.e indicates an initial state of a thing or a thing initially associated with the focused thing	ਤੋਂ	ਪਟਿਆਲਾ ਤੋਂ ਦਿਲੀ ਦੂਰ ਹੈ
16	gol	goal state i.e. indicates a final state of object or a thing finally associated with the object of an event	ਵਿੱਚ	ਰਾਮ ਪਟਿਆਲਾ ਵਿੱਚ ਰਹਿੰਦਾ ਹੈ
17	icl	indicates an upper concept or a more general concept	ਤਰਾਂ ਦਾ	ਕੁੱਤਾ ਇਕ ਤਰਾਂ ਦਾ ਜਾਨਵਰ ਹੈ
18	ins	Instrument i.e. indicates an instrument to carry out an event	ਦੇ ਨਾਲ	ਕਲਮ ਦੇ ਨਾਲ ਲਿਖੋ

Punjabi Case Markers Correspondence with UNL Relations Contd..

19	int	Intersection i.e. indicates all common instances to have with a partner concept	null	ਇਕ ਕਿਤਾਬ
20	iof	an instance of i.e. indicates a class concept that an instance belongs to	null	ਪਟਿਆਲਾ ਪੰਜਾਬ ਦਾ ਸ਼ਹਿਰ ਹੈ
21	man	manner i.e. indicates a way to carry out an event or the characteristics of a state	null	ਜਲਦੀ ਚਲੋ
22	met	method or means i.e. indicates a means to carry out an event	ਦੇ ਨਾਲ	ਪੱਥਰ ਹਥੜੇ ਦੇ ਨਾਲ ਤੋੜਿਆ ਗਿਆ
23	mod	modification i.e. indicates a thing that restricts a focused thing	ਦਾ	ਚੰਦਰਮਾ ਧੱਰਤੀ ਦਾ ਉਪਗ੍ਰਹਿ ਹੈ
24	nam	name i.e. indicates a name of a thing	null	ਓਹ ਰਾਮ ਹੈ
25	obj	affected thing i.e. indicates a thing in focus that is directly affected by an event or state	ਤੋ	ਰਾਮ ਨੇ ਸੀਤਾ ਤੋ ਕੰਮ ਕਰਾਇਆ

Punjabi Case Markers Correspondence with UNL Relations Contd..

26	opl	affected place i.e. indicates a place in focus affected by an event	ਤੋਂ	ਵਿਚਕਾਰ ਤੋਂ ਕੱਟੇ
27	or	disjunction i.e. indicates a partner to have disjunctive relation to	ਜਾਂ	ਤੁਸੀਂ ਅੱਜ ਇਥੇ ਰਹੋਗੇ ਜਾਂ ਨਹੀਂ
28	per	proportion/rate/distribution i.e. indicates a basis or unit of proportion, rate or distribution	ਵਾਰ	ਮੈਂ ਚਾਰ ਵਾਰ ਉਥੇ ਗਿਆ
29	plc	indicates a place where an event occurs, or a state that is true, or a thing that exists	ਵਿਚ	ਰਸੋਇ ਵਿੱਚ ਖਾਨਾ ਬਣ ਰਿਹਾ ਹੈ
30	plf	initial place i.e. indicates a place where an event begins or a state that becomes true	ਤੋਂ	ਪਟਿਆਲਾ ਤੋਂ ਜਪਾਨ ਤਕ
31	plt	final place i.e. indicates a place where an event ends or a state that becomes false	ਤਕ	ਪਟਿਆਲਾ ਤਕ ਸਫਰ
32	pof	part of i.e. indicate a concept of which a focused thing is a part	ਦਾ/ਦੀ/ਦੇ	ਕਿਤਾਬ ਰਾਮ ਦੀ ਹੈ

Punjabi Case Markers Correspondence with UNL Relations Contd..

33	pos	possessor i.e. indicates the possessor of a thing	null	ਮੇਰੀ ਕਿਤਾਬ
34	ptn	partner i.e. indicates an indispensable non-focused initiator of an action	ਦੇ ਨਾਲ	ਰਾਮ ਦੇ ਨਾਲ ਸਾਮ ਰਹਿੰਦਾ ਹੈ
35	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 ਸਾਮ ਝੂਠ ਬੋਲ ਰਿਹਾ ਹੈ
36	qua	quantity i.e. indicates the quantity of a thing or unit	Null	ਦੋ ਚੂਹੇ
37	rsn	reason i.e. indicates a reason why an event or a state happens	ਦੇ ਕਰਨ	ਮੀਂਹ ਦੇ ਕਾਰਨ ਸਾਰੀ ਫਸਲ ਖਰਾਬ ਹੋ ਗਈ

Punjabi Case Markers Correspondence with UNL Relations Contd..

38	scn	scene i.e. indicates a scene where an event occurs, or state is true, or a thing exists	ਵਿੱਚ	ਰਾਮ ਦੀ ਤਸਵੀਰ ਕੱਲ ਟੀ ਵੀ ਵਿੱਚ ਦਿਖਾਈ ਗਈ
39	seq	sequence i.e. Indicates a prior event or state of a focused event or state	ਦੇ ਬਾਅਦ	ਤੀਨ ਦੇ ਬਾਅਦ ਚਾਰ
40	src	indicates the initial state of an object or thing initially associated with the object of an event	ਤੋਂ	ਬਤੀ ਲਾਲ ਤੋਂ ਹਰੀ ਹੋਈ
41	tim	time i.e. indicates the time an event occurs or a state is true	ਨੂੰ	ਮੰਗਲਵਾਰ ਨੂੰ ਚਲਾਂਗੇ
42	tmf	initial time i.e indicates the time an event starts or a state becomes true	ਤੋਂ	ਸਵੇਰ ਤੋਂ ਲੈਕੇ ਸ਼ਾਮ ਤਕ

Punjabi Case Markers Correspondence with UNL Relations Contd..

43	tmt	final time i.e indicates a time an event ends or a state becomes false	ਤਕ	ਸ਼ਾਮ ਤਕ ਘਰ ਪਹੁੰਚਾਗੇ
44	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
45	via	an intermediate place or state	ਵਲੋਂ	ਪਟਿਆਲਾ ਵਲੋਂ ਜਲਦੀ ਪਹੁੰਚਾਗੇ

Use of Case Markers in Deconverter

Ram ate dinner.

{unl}

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

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

{/unl}

Output without case marker

rwm KwXw KwDw

This output is not meaningful. To generate the meaningful sentences we have to apply the case markers with in the sentences.

Rule for application of case marker

If 'agt' has parent UW with *@past* attribute and it is verb, and if child UW is noun then put case marker ny after the child UW i.e. Ram.

So ultimately the output after applying the case marker rule will be,
rwm ny KwXw KwDw

Use of Case Markers in Deconverter Contd..

Pest comes out from egg.

Partial UNL

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

Output without case marker rule

```
bcw AMfy bwhr AwauMdw hY
```

Again we can see that this sentence is meaningless, 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 qo case marker after child as egg.

After applying the case marker rule the output is:

```
bcw AMfy qo bwhr AwauMdw hY
```

Case Markers for Resolving Relations in Punjabi Enconverter

- Handling of plc Relation

+{N,CONCRETE,^CASEADD,INANI:+CASEADD:plc:}{[ਿ ਵੱ ਚ]::}

ਮੈਂ ਖੰ ਜਾ ਬਿ ਵੱ ਚ ਰਿ ਹੰ ਦਾ ਹਾੰ

+{N,CONCRETE,INANI,PLACE,^CASEADD:plc:}{V:::};

ਮੈਂ ਖੰ ਜਾ ਬਜਾ ਰਿ ਰਹਾ ਹਾੰ

Case Markers for Resolving Relations in Punjabi Enconverter Contd..

Handling of agt Relation

+{N,ANI:+CASEADD:agt:} {[ਨੇ]::}

ਮੁੰਡੇ ਨੇ ਚੈ ਲਖਾ ਯੇ ।

+{N,ANI,PERSON::agt:} {V:::}

ਰਾਮਦੇੜਚੁੱਕਿਆਹੋਵੇਗਾ ।

Conclusion

- For any language study of case constructs are very important for the generation of meaningful sentences.
- Matrix based approach is useful to design and develop language independent Deconverter.
- Sharing of rules should be there so that one can take the benefit of similar activity done on some other language.

References

- Bhattacharyya Pushpak, 2001, 'Multilingual information processing using Universal Networking Language', Indo UK Workshop on Language Engineering for South Asian Languages (LESAL), Mumbai, India.
<http://www.acl.ldc.upenn.edu/P/P00/P00-1077.pdf>
- Dave Shachi and Bhattacharyya Pushpak, 2001, 'Knowledge extraction from Hindi texts', Journal of Institution of Electronic and Telecommunication Engineers, 18(4).
<http://www.cse.iitb.ac.in/~pb/papers/iete.pdf>
- Dave Shachi, Parikh Jignashu and Bhattacharyya Pushpak, 2001, 'Interlingua based English Hindi machine translation and language divergence', Journal of Machine Translation (JMT), 16(4), pp. 251-304.

References Contd..

- Dey Kuntal and Bhattacharyya Pushpak, 2003, 'Analysis and generation of Bengali case structure in the Universal Networking Language framework', Proceedings of International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies, Alexandria, Egypt. <http://www.cfilt.iitb.ac.in/convergence03/>
- Dhanabalan T, Geetha T.V., 2003, 'UNL Deconverter for Tamil', Proceedings of International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies, Alexandria, Egypt. <http://www.cfilt.iitb.ac.in/convergence03/>
- Uchida Hiroshi, Zhu Meiyong, 2002, 'The Universal Networking Language beyond machine translation', Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Spain. http://www-clips.imag.fr/geta/User/wang-ju.tsai/articles/UNL-beyond_MT.doc
- Uchida Hiroshi, Zhu Meiyong, 2005, 'The Universal Networking Language (UNL) specifications 2005', United Nations University, Tokyo. <http://www.uncl.org/unlsys/unl/UNL2005/>

Thanks