Interdisciplinary Journal of Linguistics
Volume [10] 2017, Pp.33-39

Semantic Relations in UNL: A Study of Kashmiri Possessive Determiners

Sumaira Nabi*
Aadil Amin Kak**

Abstract

Universal Networking Language (UNL) is a computer language created to represent and process information across language barriers (Uchida et al, 2001). The primary objective of UNL is to serve as an infrastructure for handling knowledge. The present paper is an attempt to represent the semantic relation between a possessive determiner and the noun following it in the UNL framework. The UNLised text so derived will act as an important milestone in the field of Machine Translation for Kashmiri and is expected to have an F score of 1 i.e. an accuracy of almost 100%.

Key words: Universal Networking Language, UNLisation, Interactive Analyser, semantic relation etc.

1. Introduction

Universal Networking Language is a computer language created to represent and process information across language barriers (Uchida et al, 2001). UNL is basically a knowledge representation language i.e. it is used to represent information conveyed by natural languages (Cardeñosa et al, 2009). UNL represents an interpretation of a natural language text and not its translation. UNL expressions have no ambiguity as is the case with natural languages. The goal of UNL is to represent "what was meant" and not "what was said." UNL provides an infrastructure for machines to handle what is meant by natural languages, and can also be used for the purpose of translation. UNL expressions are semantically complete in order to be understandable to machines. UNL is independent of any particular language. UNL has all the components of a natural language. The two fundamental movements in UNL are:

a) UNLisation:UNLization is the process of representing the information conveyed by NL into UNL, and;
b) NLisation is the process of generating a natural language document out of UNL.

The process of UNLisation and NLisation are carried out independently by two online tools IAN (Interactive Analyser) and EUGENE (deep to surface structure Generator) provided by the UNDL Foundation. UNLisation does not take into account the format of any future NLisation and similarly NLisation does not need any information about the source language.
The information in UNL is represented through three semantically discrete entities:

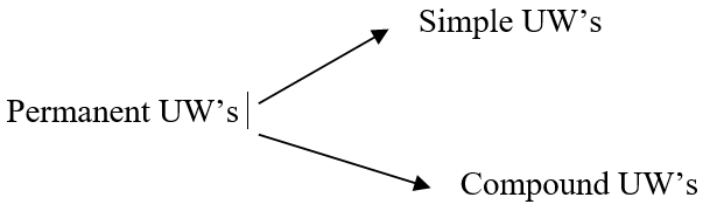
1.1. Universal words

UW's are the nodes of the UNL graph and can represent simple or compound concepts. UW's constitute the lexicon of the UNL system. UW's can be divided into two categories:

1.1.1 Temporary UW's: are the words which are not translatable and are not included in the UNL dictionary. For example, a number, an email or URL etc.

* Department of Linguistics, University of Kashmir
** Department of Linguistics, University of Kashmir

1.1.2 Permanent UW's: are the words which are conceived as single lexical items and therefore included in the UNL dictionary. For example, pen, book, mug etc.



Simple UW represents an isolated node in the UNL graph e.g. big, book, tall etc.
Compound UW represents a node combined with an attribute such as the english word "bigger" is represented as big.@more in the UNL graph.

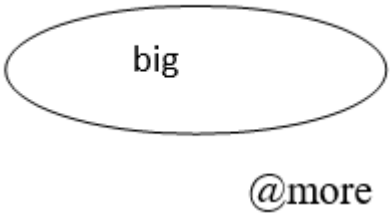
1.2 Attributes
Attribute labels express additional information about the UW's that appear in a sentence. It includes tense, number, aspect and represents information on the role of the node in the UNL graph as in the case of @entry that indicates the main node of the graph.

1.2.1 Relations
Relations formerly known as "links" are labeled arcs connecting a node to another node in a UNL graph (Martins, 2002).

2. UNL structure:
The system of UNL consists of UNL, Language servers, and basic tools. UNL involves Universal words, Relations and Attributes and Language servers involve Enconverter and Deconverter.

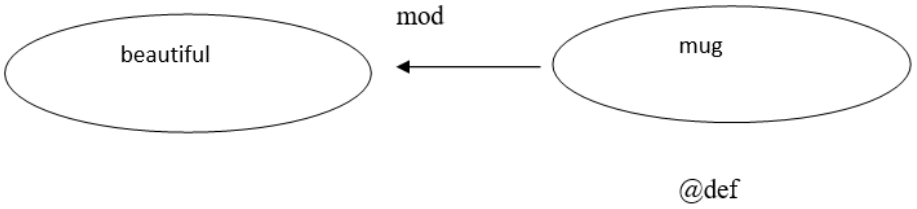
3. Graphical Representaion in the UNL system: In the UNL system, information is represented through UNL graphs.

3.1. Graphical Representation of words in the UNL System : Consider the example 'bigger'



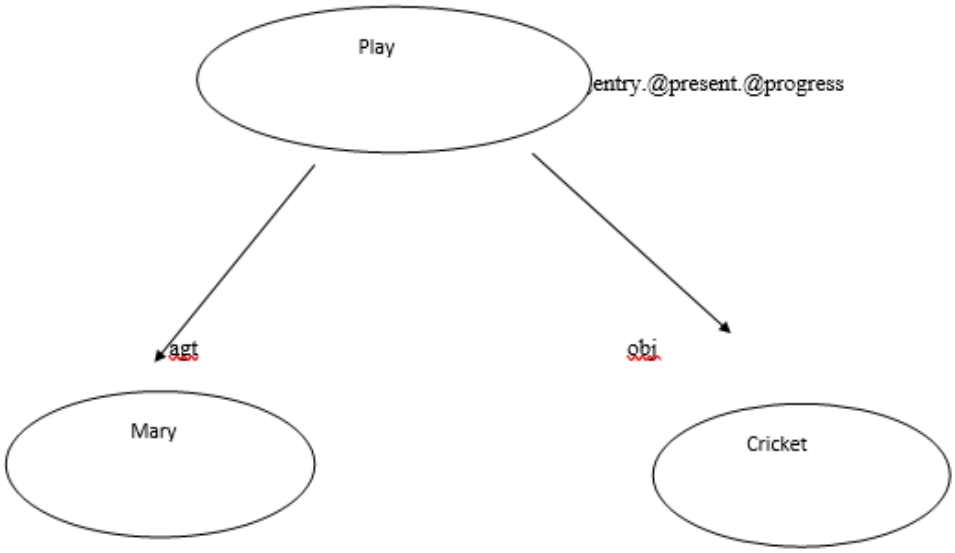
UNL Representation: big.@more
In the above graph, 'big' is a **UW** and **@more** is an attribute assigned to it.

3.2 Graphical Representation of Phrases in the UNL System: Consider the example 'the beautiful mug'



In this graph, "beautiful" and "mug" are **UWs**, **mod** (modifier) is a relation and **"@def**" is an attribute assigned to "mug".
The UNL format of the above phrase can be put as: mod (mug.@def, beautiful), This in simple terms can be interpreted as '*mug* which is definite (grammatically) and is modified by *beautiful'*, thus giving us the term 'the beautiful mug'.

3.3 Graphical Representation of Sentences in the UNL System : consider the example, 'Mary is playing cricket'



In this graph, 'play', 'Mary' and 'cricket' are **UW's .**
agt (agent) and **obj** (object) are relations and @entry.@present.@progress are attributes assigned to the **UW** ' play'.
UNL Representation: agt(play.@entry.@present.@progress, Mary)
Obj (play.@entry.@present.@progress,Cricket)

4. Methodology: The methodology for the present work involves providing linguistic Input to the Encoverter. Enconverter converts Natural Language text into UNL. The process of Enconversion requires a word dictionary and Analysis Grammar. Enconverter is a language independent software applicable for any language.

4.1 Dictionary Entries

The format of words in the IAN dictionary is as:

[HW]{ID} "UW" (ATTR….) <FLG, FRE, PRI);

Where,

HW= Head Word

ID= Identification of Head Word (omitable)

FLG= Language Flag

FRE=Frequency of Head Word

PRI= Priority of Head Word

Some examples of dictionary entries in IAN are given below:

[کِتاب]{}"book"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; a book, the book /kita:b/
[دأر]{}"window"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /dəːr/
[مگ]{}"mug"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /mag/
[کار]{}"car"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /kaːr/
[ٹرٕین]{}"train"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;
[کوٗر]{}"girl"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /kuːr/
[سِکیٖل]{}"ruler"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /sIkeːl/
[بوتَل]{}"bottle"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /
[لُڑکہٕ]{}"boy"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /ləḍkI/
[کمَل]{}"blanket"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /kamal/
[سنٛگَتر]{}"orange"(LEX=N,POS=NOU,NUM=INV)<kas,0,0>;
[کیٖل]{}"banana"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /keːl/
[شٕراک]{}"knife"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;
[دروٚاز]{}"door"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;
[نال سٕزٕن]{}"safety pin" (LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;

5. Analysis: Possessive Determiners (eg. *my*, *your*, *his*, *her*, *its*, *our*, *their*, etc.) are determiners indicating possession of someone or something referenced by the following noun. Examples of Possessive Determiners in Kashmiri include words like /meːnʲ/, /čeːnʲ/, /səːnʲ/ etc. Possessive Determiners in Kashmiri are the pronouns in genitive case agreeing with the complement eg. /meːnʲ kitaːb/, /mʲoːn kalam/, /čeːnʲ kitaːb/, čoːn kalam/

Person	Gender and Number			
	Masculine		Feminine	
	Sg.	Pl.	Sg	Pl.
1st Sg.	/mʲoːn/	/meːnʲ/	/meːnʲ/	/mʲaːni/

1st Pl.	/soːn/	/səːnʲ/	/səːnʲ/	/saːni/
2nd Sg.	/čoːn/	/čəːnʲ/	/čəːnʲ/	/čaːni/
2nd Pl.	/tuhund/	/tuhɨndʲ/	/tuhɨnz/	/tuhɨnzɨ/
3rd Sg.	/təmʲsɨnz/	/təmʲsɨnzɨ/	/humʲsənz/	/humʲsənzɨ/

In the UNL system, Possessive Determiners are represented by "00" followed by person attributes:

00.@1 (first person singular)
00.@1.@pl (first person plural)
00.@2 (2nd person singular)
00.@2. @ Pl (2nd person plural)
00.@3 (3rd person singular)
00.@3.@pl (3rd person plural)
00.@3.@female (3rd person singular female)
00.@3.@male (3rd person singular male)

5.1 Representation of Possessive Determiners and their Complement : The representation of Possessive Determiners in the UNL programme is shown in a tabular form below:

Result	**Dictionary**	**Rules**	**English**	**Corpus**
Pos (book,00.@1)	[**میٲنی**]{}"00.@1"(LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,255,0>; [**کتاب**]{}"book"(LEX=N, POS=NOU) <K, 0,0>;	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	my book	میٲنی کِتاب /meːnʲ kitaːb/
Pos (window,00.@2)	[چاٲنی]{}" 00.@2"(LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,255,0>; [**دأر**]{}"window"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>; /dəːr/	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	your window	چاٲنی **دأر** /čəːnʲ dəːr /
Pos (mug,00.@3@female)	[**بُمی سِنْز**]{} "00.@3@female" (LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,0,0>; [**مگ**]{}"mug"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	her mug	بُمی سِنْز کِتاب /humʲ sınz mag//
Pos (book,00.@3@male)	[**تُمی سِنْز**]{} "00.@3@male"(LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,0,0>;	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	his book	تُمی سِنْز کِتاب /təmʲ sınz kitaːb/

le)	[**کتاب**] {}"book"(LEX=N, POS=NOU) <K, 0,0>;			
Pos (blanket,00.@3)	[**أمیسِنْز**] {} "00.@3" (LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,0,0>; [**کمَل**]{}"blanket"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	its blanket	أمیسِنْز کِتاب /əmʲsınz kamal/ /
Pos (book,00.@1@pl)	[سأنی]{} "00.@1@pl" (LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,0,0>; [**کتاب**] {}"book"(LEX=N,POS=NOU) <K, 0,0>;	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	our book	سأنی کِتاب /səːnʲ kitaːb/
Pos (banana,00.@3@pl)	[**تِہِنْز**]{} "00.@3@pl"(LEX=D,POS=POD,CAS=GEN,rln=pos)<kas,0,0>; [**کیٖل**]{}"banana"(LEX=N,POS=NOU,NUM=SNG)<kas,0,0>;	(%a,D,POS=POD,CAS=GEN,rln=pos)(%b,N,POS=NOU,NUM=SNG):=pos(%b;-GEN,%a);	their book	تِہِنْز کِتاب /tihinz keːl/

6.Conclusion

The aim of this paper was to represent the semantic relation between possessive determiners and their complements in the UNL framework. The output derived in the form of semantic relations can serve as an important milestone in the development of a UNL based Machine Translation system for Kashmiri. This work is however a preliminary study leaving scope for more research in this area.

References

Bhattacharyya, Pushpak. Multilingual information processing through Universal Networking Language. *Indo UK Workshop on Language Engineering for South Asian Languages*, 2001.

Cardeñosa, Jesús, et al., (ed) 2005. *Universal Networking Language: Advances in Theory and Applications.* Instituto Politécnico Nacional Centro de Investigación en Computación.

Cardeñosa, Jesús, et al. Multilingual Cross Language Information Retrieval A new approach. *Computer Science and Information Technologies*, 2009, pp. 284-287. www.vai.dia.fi.upm.es/publicaciones/multi_draft_csit09.pdf

Prashanth, Kumar. *Semantics Extraction from Text.* 2010. IIT Bombay, M. Tech dissertation. www.cse.iitb.ac.in/alumni/~pkamle08/website/docs/thesis.pdf

Kak, Aadil, et al. 2015. Phrase Structure in Kashmiri: A UNL Approach. *Proceedings of COMMUNE*.
Nabi, Sumaira, and Sumaya Jehangir. 2015. Analysis Grammar in UNL: An Initiative towards MT for Kashmiri. *Interdisciplinary Journal of Linguistics*, vol. 8, pp. 189- 197.
Uchida, H., and Meiying Zhu. 2001. The universal networking language beyond machine translation. *International Symposium on Language in Cyberspace, Seoul*.
Uchida, H., et al. 2005. Universal Networking Language. *UNDL Foundation*.

❐❐❐