

Overview

- Introduction to the nature of syntactic representations. (Rambow, 15 minutes)
- Introduction to the morphology, syntax, and lexical semantics of Hindi and Urdu. (Sharma, 40 minutes)
- The morphological representation for Hindi and Urdu, including encoding issues, tokenization, part-of-speech tags, and morphological representation. (Sharma and Rambow, 20 minutes)
- The dependency representation (DS) for Hindi and Urdu syntax: principles, representation, and examples. (Sharma, 25 minutes)
- **The lexical semantic representation (PB) for Hindi and Urdu: principles, representation, and examples. (Vaidya, 25 minutes)**
- The phrase structure representation (PS) for Hindi and Urdu syntax: principles, representation, and examples. (Rambow, 25 minutes)
- Sample initial experiments in Hindi and Urdu NLP using the HUTB. (Sharma and Rambow, 15 minutes).

Lexical Semantic Representation for Hindi & Urdu : principles, representation and analysis

Ashwini Vaidya

University of Colorado, Boulder

Contents

1. Motivation
2. Introducing PropBank
3. Frame file definition
4. Hindi PropBank
5. Linguistic Phenomena

Why is semantic information important?

- Imagine an automatic question answering system
- Who created the first effective polio vaccine?
- Two possible choices:
 - Becton Dickinson created the first disposable syringe for use with the mass administration of the first effective polio vaccine
 - The first effective polio vaccine was created in 1952 by Jonas Salk at the University of Pittsburgh

Word Matches

- Who **created** the **first effective polio vaccine**?
 - Becton Dickinson **created** the first disposable syringe for use with the mass administration of the **first effective polio vaccine**
 - The **first effective polio vaccine** was **created** in 1952 by Jonas Salk at the University of Pittsburgh

Parsing

- Who created the first effective polio vaccine?
 - [Becton Dickinson] created the [first disposable syringe] for use with the mass administration of the first effective polio vaccine
 - [The first effective polio vaccine] was created in 1952 by [Jonas Salk] at the University of Pittsburgh

Semantic Role labelling

- Who created the first effective polio vaccine?
 - [Becton Dickinson_{agent}] created the [first disposable syringe_{theme}] for use with the mass administration of the first effective polio vaccine
 - [The first effective polio vaccine_{theme}] was created in 1952 by [Jonas Salk_{agent}] at the University of Pittsburgh

SRL gives us the right answer

- We need semantic information to prefer the right answer
- The theme of create should be 'the first effective polio vaccine'
- The theme in the first sentence was 'the first disposable syringe'
- We can filter out the wrong answer

We need semantic information

- To find out about events and their participants
- To capture semantic information across syntactic variation

We need semantic information

- To find out about events and their participants
- To capture semantic information across syntactic variation

Semantic information

- Semantic information about verbs and participants expressed through semantic roles
- Agent, Experiencer, Theme, Result etc.
- However, difficult to have a standard set of thematic roles

Proposition Bank

- Proposition Bank (PropBank) provides a way to carry out general purpose Semantic role labelling
- A PropBank is a large **annotated** corpus of predicate-argument information
- A set of semantic roles is defined for *each* verb
- A syntactically parsed corpus is then tagged with verb-specific semantic role information

PropBank Frame files

- PropBank defines semantic roles on a verb-by-verb basis
- This is defined in a verb lexicon consisting of *frame files*
- Each predicate will have a set of roles associated with a distinct usage
- A polysemous predicate can have several rolesets within its frame file

An example

- John rings the bell

ring.01	Make sound of bell
Arg0	Causer of ringing
Arg1	Thing rung
Arg2	Ring for

An example

- John rings the bell
- Tall aspen trees ring the lake

ring.01	Make sound of bell
Arg0	Causer of ringing
Arg1	Thing rung
Arg2	Ring for

ring.02	To surround
Arg1	Surrounding entity
Arg2	Surrounded entity

An example

- [John] rings [the bell] ← Ring.01
- [Tall aspen trees] ring [the lake] ← Ring.02

ring.01	Make sound of bell
Arg0	Causer of ringing
Arg1	Thing rung
Arg2	Ring for

ring.02	To surround
Arg1	Surrounding entity
Arg2	Surrounded entity

An example

- [John_{ARG0}] rings [the bell_{ARG1}] ← Ring.01
- [Tall aspen trees_{ARG1}] ring [the lake_{ARG2}] ← Ring.02

ring.01	Make sound of bell
Arg0	Causer of ringing
Arg1	Thing rung
Arg2	Ring for

ring.02	To surround
Arg1	Surrounding entity
Arg2	Surrounded entity

Hindi PropBank

- Annotating Hindi PropBank involves three steps:
 - Creation of frame files
 - Empty argument insertion
 - Semantic role labelling

Frame files for Hindi

- Two types of frame files:
 - Frames for simple verbs [385 frames; 703 predicates]
 - Frames for nominals in complex predicates [1875; 1902 predicates]

Empty Arguments

- PropBank inserts 4 types of empty arguments
 - **pro** :dropped null arguments; recoverable from discourse context
 - **PRO**: empty subjects of non-finite complement and adjunct clauses
 - **RELPRO**: gaps in participial relative clauses
 - **GAP**: elided arguments in co-ordinated clauses
- **PRO** and **RELPRO** are inserted automatically
- **GAP** and **pro** are inserted manually

PropBank Tagset

Numbered Arguments	Numbered Arguments with function tags
ARGA: Causer	ARGA-MNS: Indirect causer
ARG0: Agent, experiencer	ARG0-MNS: Induced causer
ARG1: Theme, patient	ARG0-GOL: Causee with a 'recipient' role
ARG2: Recipient	ARG2-ATR: Attribute
ARG3: Instrument	ARG2-GOL: Goal
	ARG2-SOU: Source
	ARG2-LOC: Location
	ARG2-DIR: Direction

PropBank Tagset

Modifier Arguments

ARGM-TMP : Temporal

ARGM-MNR : Manner

ARGM-LOC : Location

ARGM-PRP: Purpose

ARGM-CAU : Cause

ARGM-DIS : Discourse

ARGM-ADV : Adverb

ARGM-MNS : Means

Linguistic phenomena

- Simple transitive
- Unaccusative and Unergative
- Existential
- Dative subject
- Ditransitive
- Causatives
- Complex Predicates

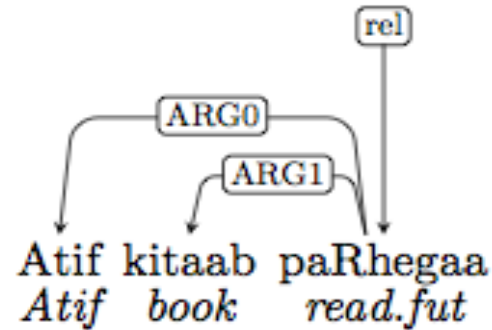
Simple Transitive

trans-1: आतिफ़ किताब पढ़ेगा

Atif kitab paRhegaa

Atif book.f read.m.sg.fut

'Atif will read the book'

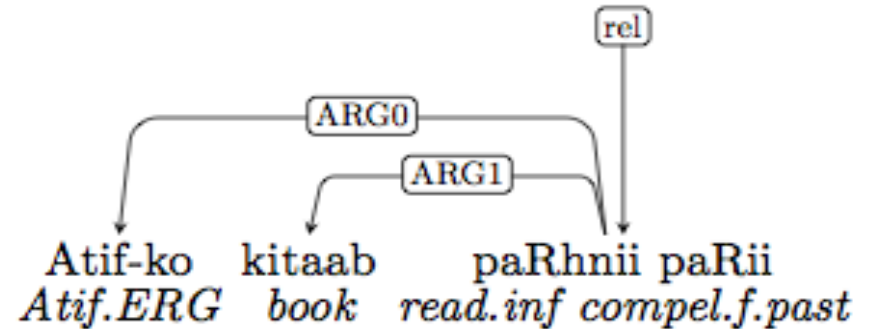


trans-2: आतिफ़ ने किताब पढ़ी

Atif ne kitaab paRhii

Atif erg book.f read.f.sg.pst

'Atif read the book'



trans-3: आतिफ़ को किताब पढ़नी पड़ी

Atif ko kitaab paRhunii paRii

Atif dat book.f read.f.inf compel.f.pst

'Atif had to read the book'

Unaccusative & Unergative

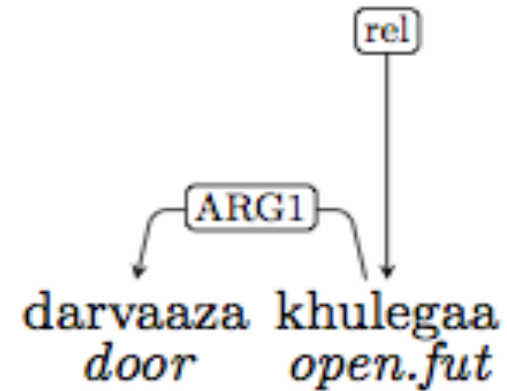
- Distinction between intransitive verbs:
 - unaccusatives e.g. Kula (open), Puta (explode)
 - Unergatives e.g. haMsa (laugh),
- Single argument of unaccusatives takes Arg1, unergatives take Arg0
- Diagnostic tests are used to distinguish unaccusatives from unergatives
 - E.g. animacy test, cognate object test among others

Intransitive: Unaccusative

unacc-1: दरवाज़ा खुलेगा
darvaazaa khulegaa
door.m.sg.d open.m.sg.fut
'The door will open'

unacc-2: *दरवाज़े ने खुला
*darvaaze ne khulaa
door.m.sg.obl erg open.pst
'The door opened'

unacc-3: दरवाज़े को खुलना पड़ेगा
darvaaze ko khulnaa paRegaa
door.m.sg.obl dat open.inf compel.fut
'The door will have to open'

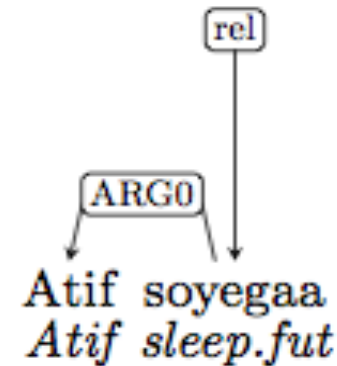


Intransitive: Unergative

Unerg-1: आतिफ़ सोएगा
Atif soyegaa
Atif sleep.m.sg.fut
'Atif will sleep'

unerg-2: *आतिफ़ ने सोया
*Atif ne soyaa
Atif erg sleep.m.sg.pst
'Atif slept'

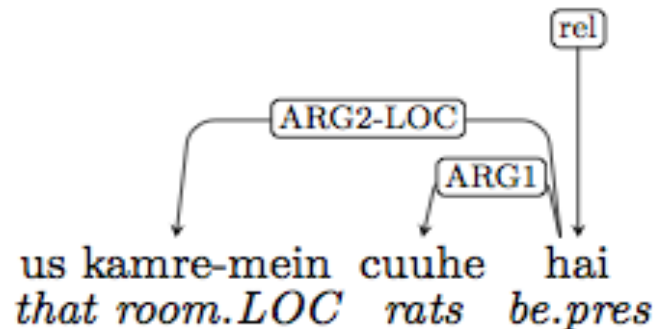
unerg-3: आतिफ़ को सोना पड़ेगा
Atif ko sonaa paRegaa
Atif dat sleep.inf compel.fut
'Atif will have to sleep'



Existential

exist-1: उस कमरे में चूहे हैं
us kamre meM cuuhe haiM
that room in rats be.pres.pl
'There are rats in that room'

- We distinguish between existential and copula sentence types by means of different roleset IDs



Dative Subject

unacc-4: कल रात बादलों में चाँद दिखा

kal raat baadaloM meiM caaMd dikhaa

yesterday night clouds in moon see(unacc).pst

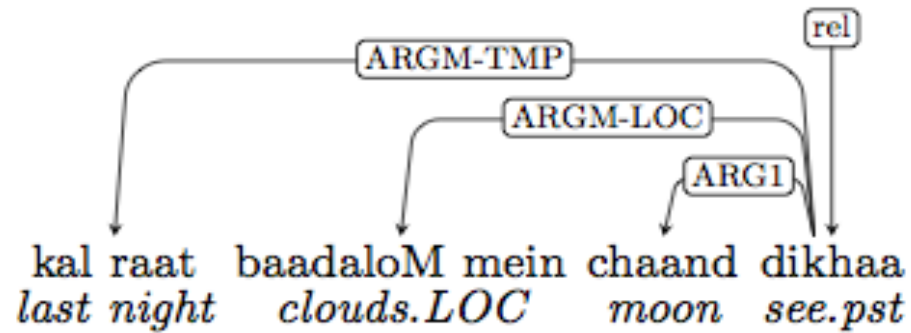
'Yesterday night, the moon was seen behind the clouds'

dat-subj-1: कल रात बादलों में मुझको चाँद दिखा

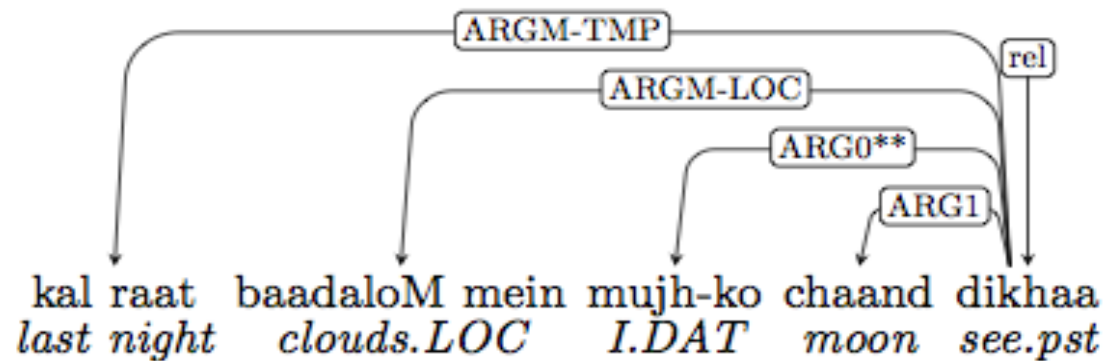
kal raat baadaloM meiM mujhko caaMd dikhaa

yesterday night clouds in me.dat moon see(unacc).pst

'Yesterday night, I saw the moon behind the clouds'



unacc-4:



dat-subj-1:

**The ARG0 analysis of dative subjects may change in future PB annotation

Ditransitive

ditrans-1: राम मोहन को किताब देगा

raam mohan ko kitaab degaa

Ram Mohan dat book.f give.m.sg.fut

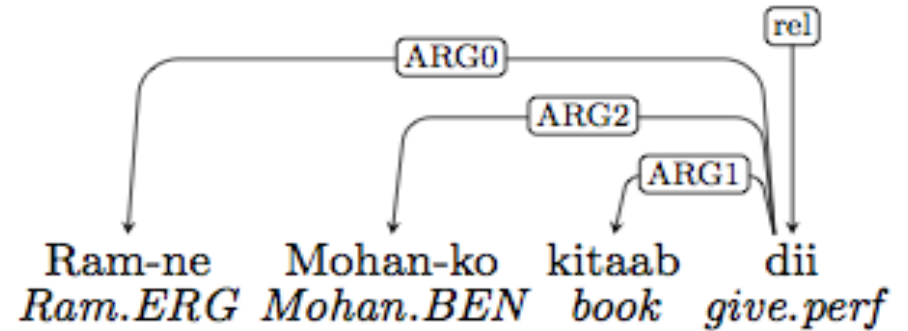
'Ram gave a book to Mohan'

ditrans-2: राम ने मोहन को किताब दी

raam ne mohan ko kitaab dii

Ram erg Mohan dat book.f give.f.sg.pst

'Ram gave a book to Mohan'



Causatives

- Hindi has two ways of forming the causative:
 - Add *-aa*
 - (so → *sulaa*) sleep → make someone sleep
 - Add *-vaa*
 - (*sulaa* → *sulvaa*) make someone sleep → cause someone to fall asleep
- We introduce the label ARGGA to analyze causers
- Subtypes of ARG0 (ARG0-GOL, ARG0-MNS) for causees
- ARGGA-MNS for intermediate causers

Causatives

Unerg-1: आतिफ़ सोएगा

Atif soyegaa

Atif sleep.m.sg.fut

'Atif will sleep'

causative-1: आया ने आतिफ़ को सुलाया

aayaa ne Atif ko sulaayaa

maid erg Atif acc sleep.caus.pst

'The maid caused the child to sleep'

causative-2: माँ ने आया से आतिफ़ को सुलवाया

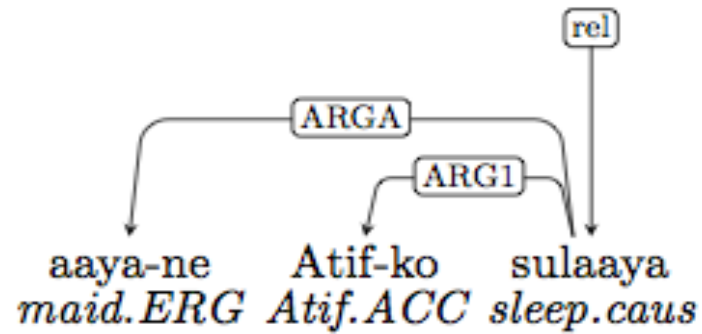
maan ne aayaa se Atif ko sulvaayaa

mother erg maid by Atif acc sleep.caus.pst

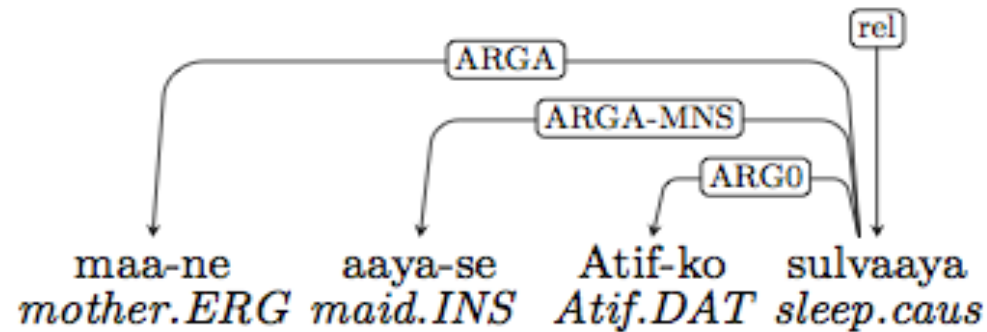
'The mother made the maid to cause the child to sleep.'

Causatives

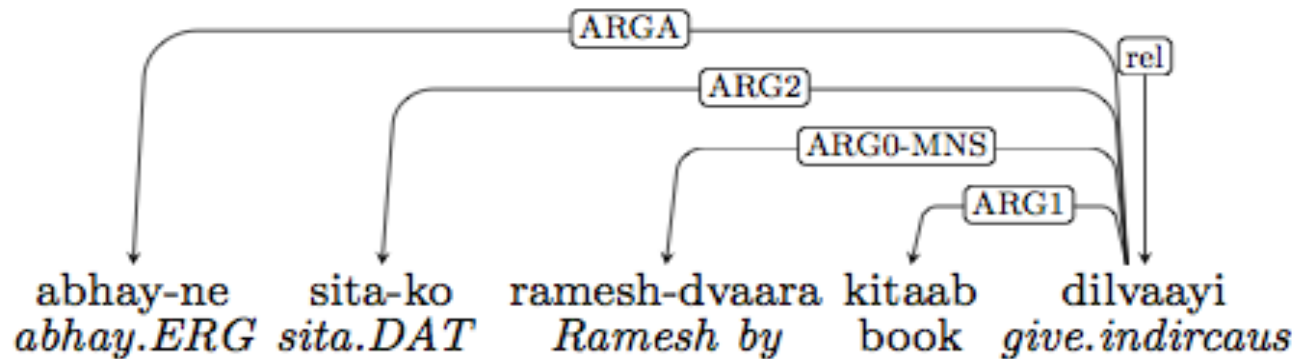
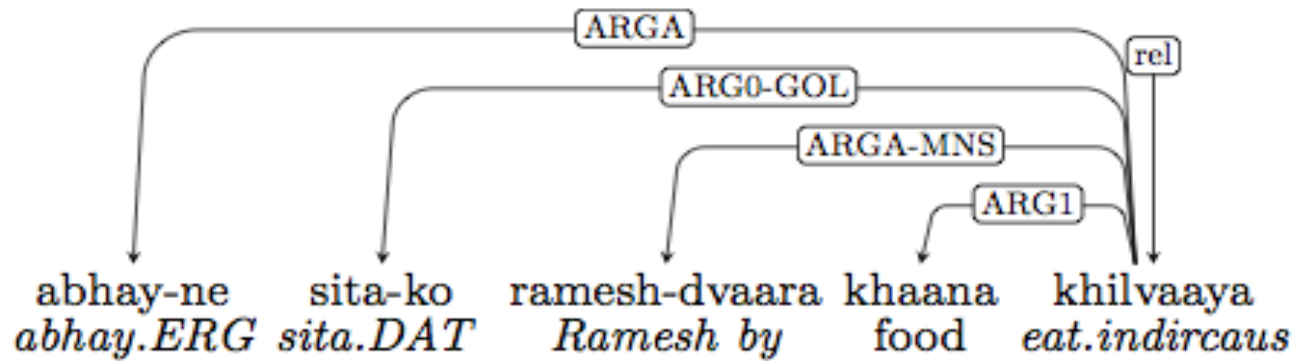
Causative-1



Causative-2



Causatives: classes



Complex predicates

- These are cases such as *bharosaa karna* `trust(n) do(v)'; trust
- Such cases are handled using a noun frame for *bharosaa*

Frame file for Barosa-n	
Barosa.01: trust-n	<i>light verb 'do' to trust</i>
Arg0	person who trusts
Arg1	thing trusted
Barosa.02 : trust-n	<i>light verb 'give' to ensure</i>
Arg0	person who ensures
Arg2	recipient of the thing ensured
Arg1	thing that is ensured

[abhay ne_{ARG0}] [sitaa par_{ARG1}] bharosaa kiya

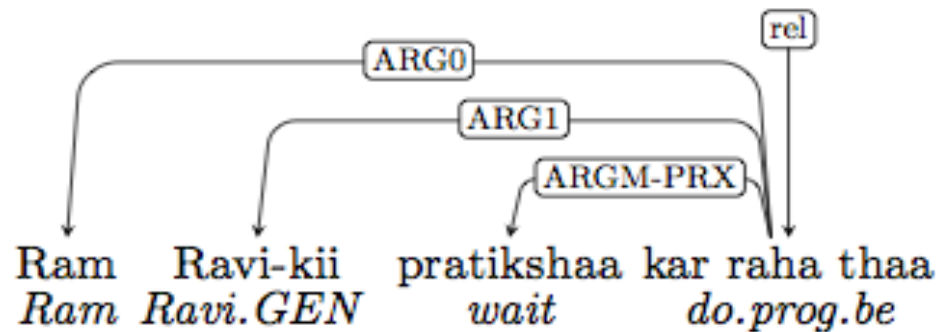
Complex Predicate

compl-pred-1: राम रवि की प्रतीक्षा कर रहा था

raam ravi kii *pratikshaa* kar rahaa thaa

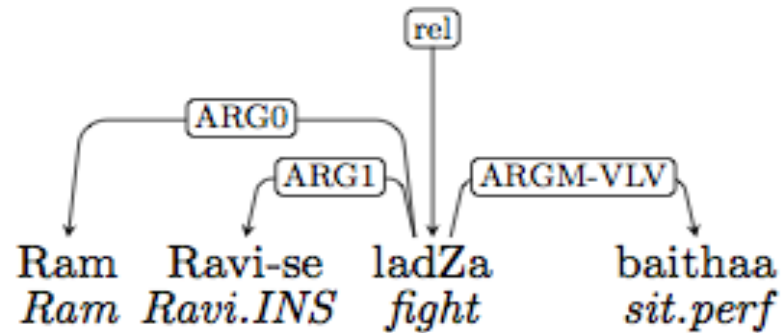
Ram Ravi gen wait do prog.m.sg be.m.sg.pst

'Ram was waiting for Ravi'



Complex predicate

compl-pred-2: राम रवी से लड़ बैठा
raam ravi se *ladZa baithaa*
Ram Ravi inst fight *sit.perf*
'Ram regrettably fought with Ravi'



Complement Clause

compl-cl-1: राम जानता है कि सीता देर से आएगी

raam jaantaa hai ki siita der se aayegii

Ram know.hab.m.sg be.sg.pres that Sita late part come.f.sg.fut

'Ram knows that Sita will arrive late'

