Paninian Computational Dependency Grammar

Dipti Misra Sharma
LTRC, IIIT, Hyderabad
Lecture CL-1 : April 4, 2025

Dependency Grammar

- Modern dependency grammar proposed by Lucien Tesnière, a French linguist (1959)

- In a dependency grammar
  - All words in a sentence depend on another word except one word
    - The word which does not depend on any other word is the 'root' (normally a verb)

Some Dependency Grammar Schemes

- Functional Generative Description (FGD)
  - Prague dependency treebank
- Stanford dependency framework
  - Stanford parser
- Universal dependency grammar framework
- Computational Paninian grammatical framework
  - Hindi/Urdu dependency treebanks

In the Indian grammatical tradition, Paninian grammar already had a grammar formalism which is dependency grammar based.

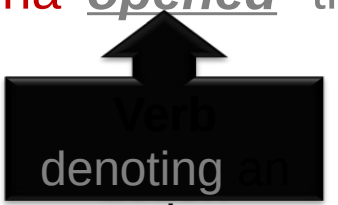
Panini's Grammar

- Dated around 500 B.C.
- Seeks to provide a complete, maximally concise and theoretically consistent analysis of Sanskrit grammatical structure
- Based on spoken form <Kiparsky, 1993>

- Focuses on language as a means of communication

Relations between words

Yesterday Sabina ***opened*** the lock with this key at my home



Participants

Panini's Grammar contd

- Treats a sentence as a series of modifier-modified relations
- Every sentence has a primary modified ('root ' generally a verb)
- Relations between verbs and their participants called ‘karaka’
- Other relations – such as reason, prupose, genitive etc
- The relations are expressed through explicit markers called 'vibhakti'

Relation between words : Paninian Perspective

Paninian Grammar has the notion of *Karaka*,

- *karakas* are the relations between a verb and its arguments
- *karakas* are the direct participants of the action denoted by a verb

k1 (kartaa) : the doer of the action
k2 (karma) : The most desired by the doer (patient/theme)
k3 (karaNa) : instrument
k7p (deshaadhikaraNa) : place
k7t (kaalaadhikaraNa) : time

Karaka Relations

- Direct participants in an action/event
- Syntactico-semantic
- Total six karaka relations
  - *kartaa*, *karma*, *karaNa*, *sampradaan*, *apaadaan*, *adhikaraNa*
-
- *kartaa* and *karma* of a verb are determined by the verb's semantics
- Verb denotes an action/event
- Any action is a bundle of sub-actions

Sabina opened the lock with this key
This key opened the lock
This lock opened

Grammatical Roles, Theta Roles, Karaka Roles

Example	Gram Role	Theta Role	Karaka Role
Peter Closed the door	Subject/ Object	Agent/ Theme	Kartaa/ Karma
The door closed	Subject	Theme	karta
The door was opened by Peter	Subject/ by phrase	Theme/ agent	Karma/ kartaa

Grammatical Roles, Theta Roles, Karaka Roles

Example	Gram Role	Theta Role	Karaka Role
Peter Closed the door	Subject/ Object	Agent/ Theme	Karta/ Karma
The door closed	Subject	Theme	karta
The door was opened by Peter	Subject/ by phrase	Theme/ agent	Karma/ karta

	Theta role	Karakas
	Agent	Kartaa/ karma
	theme	Kartaa/ karma
	instrument	karaNa
	source	apaadaana
	Receipient	sampradaan
	location	adhikaraNa
	Force	----

Action denoted by a verb: A bundle of sub-actions

Opening (of lock)

Inserting and turning a key (Action1)

Sabina **opened** the *lock*

open

k1 k2

Sabina *lock*

Key pressing against and moving lever (Action 2)

The Key **opened** the *lock*

open

k1 k2

Key *lock*

karaNa -> kartaa

Latch moving and lock opening (Action 3)

The *lock* **opened**

open (unaccusative)

k1

lock

karma -> kartaa

Languages may encode this information differently

- Some may mark the difference morphologically. (??)
- English, in most cases, has the same verb. (eg. open)
- Some languages may encode different sub-actions in different verb roots.
  - Hindi, in most cases, has a different verb for unaccusatives. (khol, khul)

Opening of lock

Inserting and turning a key (Action1)

Sabina ne taalaa kholaa
'Sabina **opened** the lock'

kholaa
(open)

k1 k2

Sabina *taalaa*

Key pressing against and moving lever (Action 2)

Is chaabi ne taalaa kholaa
'This key **opened** the lock'

kholaa
(open)

k1 k2

chabii *taalaa*

karaNa-kartari

Latch moving and lock opening (Action 3)

taalaa khulaa (unaccusative)
The lock **opened**

khulaa
(open)

k1

taalaa

Thus

- An action is a bundle of sub-actions
- Languages may encode the sub-actions differently
- A speaker may choose to express/focus on any sub-action of a larger action
- Each sub-action has its own *karta*
- The *karta* in a sentence is realised depending on the chosen sub-action
- What and how the speaker expresses an event is ‘vivaksha’ (speaker’s intention)
- Syntax reflects ‘vivaksha’ (Bharati et al 1995)

Semantics of the verb

- A verbal root denotes:
  - The activity
  - The result

Verbal Root

activity **result**

- Locus of activity : *karta*
- Locus of result : *karma*

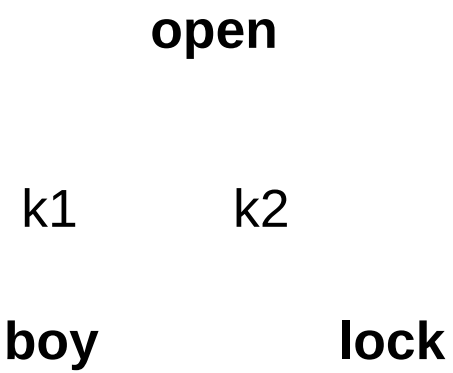
karta - karma

- The boy opened the lock
  - k1 – *karta*
  - k2 – *karma*

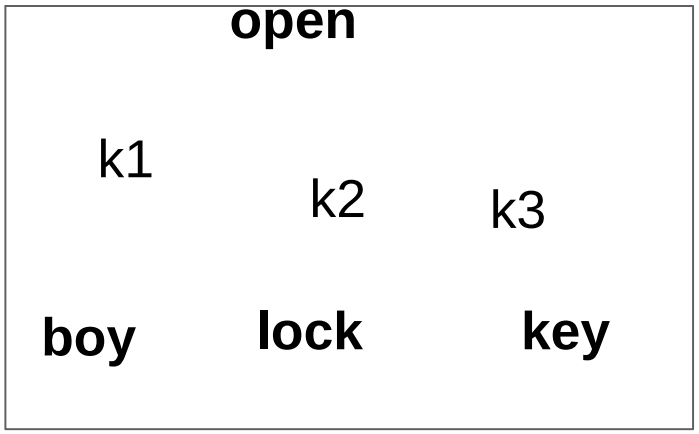
- *karta*, *karma* sometimes correspond to agent/theme
  - Not always

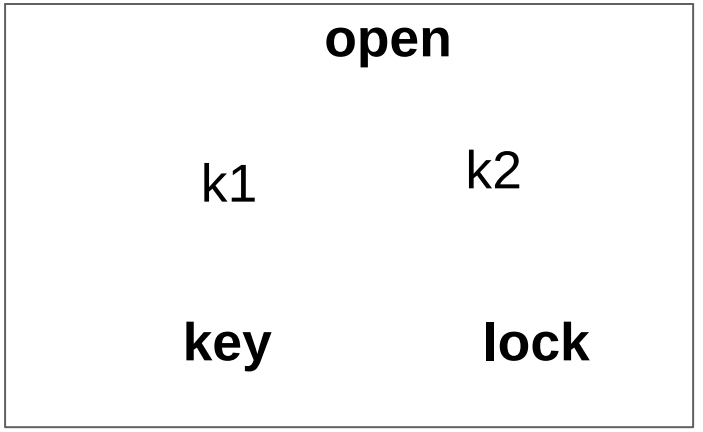
The door opened

- 'The door' is karta
- The sentence has no explicit karma

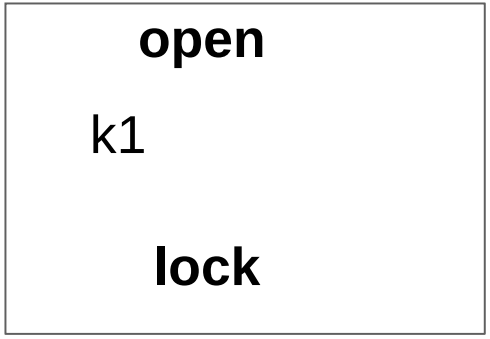


Sub-actions - Opening of lock





k1 – *karta (doer)*
k2 – *karma (affected)*
k3 – *karana (instrument)*



Vivaksha
(Speaker’s intention)

- The action of 'opening' normally requires an agentive participant. So,

Sabina opened the lock

However,
- The speaker may decide not to express the role of the agent. Hence,

The key opened the lock

- 'karaNa' is raised to the role of 'karta' (karaNa-kartri).
- The action expressed is that of the ‘karaNa’.

The lock opened

'karma' is raised to the role of 'karta' (karma-kartri)
The action expressed is of the ‘karma’

Thus, karta or other karaka roles can 'shift' depending on what the speaker wants to express (vivaksha)

Speaker’s Intention *(vivakshaa)*

- Every sentence reflects speaker’s intention
  - Participants are assigned various relations accordingly

*(a) 'I opened the lock with **this key**'*

*(b) 'I am sure **this key** will open the lock'*

  - ‘key’ gets assigned *karta (in b)*, *karana (in a)* based on what the speaker wants to express

- Syntax reflects *vivaksha*

Aakaankshaa
(A Word's Semantic Expectation)

Every word has certain Aakaankshaa (demands) to be fulfilled for the meaning to be completed for the hearer.

- A word uttered in isolation raises certain expectations.
- If a noun is uttered alone
  - 'Sabina' ???
  - Expectation 'what did she do?/what happened to her?/where is she?/etc? (an action/event/state)
- If a verb is uttered alone
  - 'Left' ???
    - Expectation : 'who left?', 'when?', 'etc?'

Verb
(Its Syntactic and Semantic Requirements)

- A verb plays a crucial role in the scheme
- Verb frames are a major help

Also,

- The arguments of a verb need to satisfy certain semantic conditions (yogyataa)

Yogyataa (Eligibility)

Selectional Restrictions

For example,

baccaa phala khaataa hai
Child fruit eat-hab pres

- *'phala' (fruit) does not have the eligibility to become the 'karta' of the verb 'khaa' (eat)*

- *Constraints based on **yogyata** require semantic knowledge for each lexical item*

This knowledge can either be annotated or can be obtained from an existing lexical resource such as a 'WordNet'

Sannidhi (Proximity)

- The modifier and the modified tend to occur in close proximity in a sentence

For example,
'Ram ne kelaa khaayaa, Mohan ne duudha piyaa Ora Hari ne film dekhii'
Ram erg banana ate , Mohan erg milk drank and Hari erg movie saw

- This Hindi example contains three verbs -

khAyA (ate), *piyA* (drank) and *dekhI (saw)*

Respective arguments of each of these verbs would tend to occur in close proximity to them

Relations in CPG

- Karaka relations
- Non-karaka relations
- Non-dependency relations

Realization of *karaka* : *vibhakti* (relation markers) (1/2)

- *karakas* are semantic relations
- They are realized in a sentence by a ‘vibhakti’

(relation marker/case marker)

- Nouns are inflected for ‘vibhakti’ in morphologically rich languages. (Sanskrit, Telugu, Tamil etc)
- Verbs are inflected for their tense, aspect, modality
- In some languages, verbs are also inflected for number, gender and person

Thus, ‘words’ are composed of, at least, two parts

Realization of *karaka* : *vibhakti* (relation markers) (2/2)

An example from Telugu, an Indian language

బాలుడితో
(‘baluDito’)

Realization of *karaka* : *vibhakti* (relation markers) (2/2)

An example from Telugu, an Indian language

బాలుడితో (with boy)
(‘baluDito’)

Realization of *karaka* : *vibhakti* (relation markers) (2/2)

An example from Telugu, an Indian language

బాలుడితో
(‘baluDito’)

బాలుడి
boy

తో
with

Similarly:- బాలుడికి (‘baluDiki’, to boy)
తోటలో (‘toTalo’, in garden)

Thus, ‘words’ are composed of
root (content morpheme) : *prakriti*
suffix (relation marker): *pratyaya*

Similarly

Take the English verb

opened

open
prakriti
verb root

ed
pratyaya
suffix

- Grammatically inflected word forms in Paninian grammar are called '*pada*'
- Only a *pada* can participate in sentence formation
- A *pada* encodes the information of what role it is playing in a sentence and what relation it holds to another word in a sentence

Pada

Panini's sutra *suptiṅantaṁ padam* states:

- prātipadika + sup = subanta pada
  nominal base + nominal inflection = nominal pada

- dhātu + tiṅ = tiṅnta pada
  verbal base + verbal inflection (finite) = verbal pada

Therefore,
Nominal and verbal inflections in a word provide syntactic cues

What about languages such as English and Hindi?

Which do not mark most of this information at the word level. Example,

Eng : Sabina was opening the lock with this key

Hin : *Sabina chaabii se taalaa khol rahii thii*
'Sabina' 'key' 'with' 'lock' 'open' 'PROG' 'past'

Eng : Sabina should open the lock with this key

Hin : *Sabina ko is chaabii se taalaa kholnaa chaahiye*
'Sabina' 'dat' 'this' 'key' 'with' 'lock' 'open-inf' 'should'

Akshar Bharati's Solution (for Hindi)

Sabina is chabii_se taalaa khola_rahii _thii
'Sabina' 'this' 'key_with' 'lock' 'open_prog_past'

Sabina ko is chaabii_se taalaa kholnaa_chaahiye
'Sabina' 'dat' 'this' 'key_with' 'lock' 'open-inf_should'

- Compute postpositions (vibhakti) and TAM (auxiliary verbs) and group them with nouns and verbs respectively (LWG)
- Get a unit equivalent to a 'pada'
- Use postpositions and auxiliaries as morphological features for ML/rules etc.

Where is the information encoded (in Hindi/English)?

- Prakriti (root)
- Vibhakti (pratyaya) : nominal and verbal inflections
- Agreement features such as number , gender , person etc.
- In English like languages, it can also be in position

ve is chaabii se taalaa khol rahe haiM
‘They are opening the lock with this key’

khol _’ **rahe_haiM** ‘are_opening’

karta	karana	karma
0	se ‘with’	o
ve ‘they’	chaabii ‘key’	taalaa ‘lock’

Sabina ne is chaabii se taalaa kholaa
‘Sabina opened the lock with this key’

khol aa ‘opened’

karta
ne ‘erg’

karana

karma

Sabina

chaabii ‘key’

taalaa ‘lock’

Sabina ko is chaabii se taalaa kholnaa paDZaa
‘Sabina had to open the lock with this key’

khol naa_paDZaa ‘had_to_open’

karta	karana	karma
ko		
Sabina	Chaabii ‘key’	taalaa ‘lock’

Relations are expressed

Through *vibhaktis*

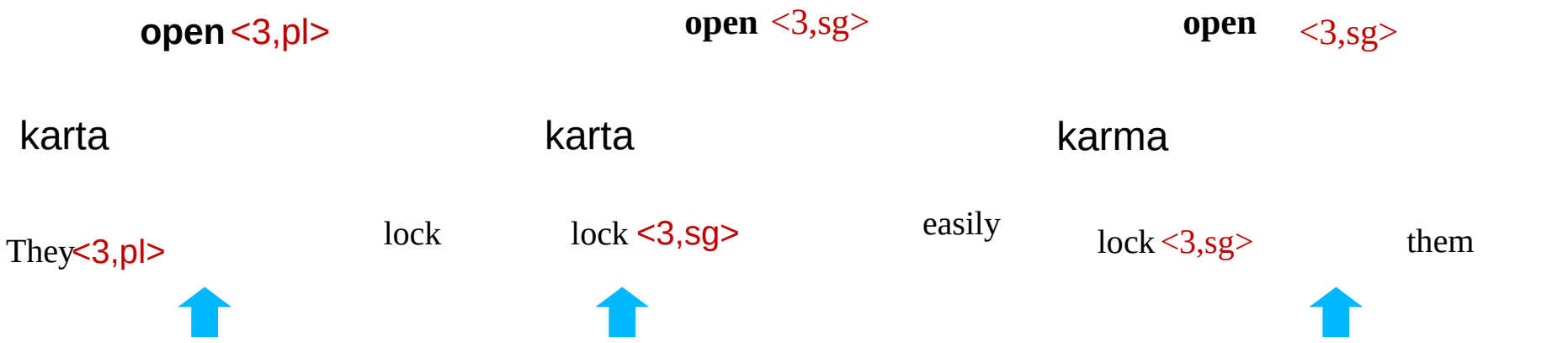
- Nominal inflections (case markers)

However, one of the relations (in some languages) is expressed by the verbal inflection through agreement (abhihit/ukta)

- The noun in this situation occurs in nominative case

Abhihita (Expressed *karaka*)

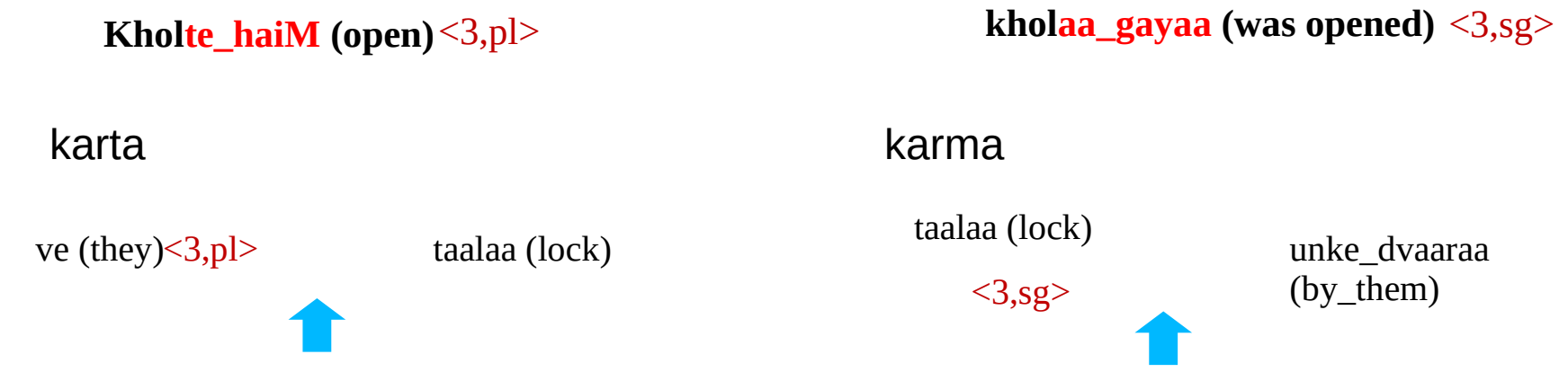
They open the lock
The lock opens easily
The lock was opened by them



Abhihita (Expressed *karaka*)

ve taalaa kholte haiM (they open the lock)

taalaa unke dvaaraa kholaa gayaa (the lock was opened by them)



Active voice (kartari prayoga), Passive voice (karmaNi prayoga)

Relations are marked

Generally relations are marked in three ways in a language:

1. Through position (Subject in English)
2. Through suffixes (morpho-syntax)
3. Through meaning of the words and their eligibility for fulfilling a particular role (‘yogyata’ in Paninian term)

The basic building block in a sentence is a ‘pada’

However

No one-to-one correspondence between relations and relation markers

1. Manish kitab padh raha hai
2. Manish ne kitab padhi
3. Manish ko kitab padhani chahiye
4. Manish se kitab nahin padhi jaa rahii

‘karta’ in all the above sentences has a different ‘vibhakti’ - 0, ne, ko and se.

Syntactic Cues

- Verbal inflections (Tense Aspect Modality (TAM))
  - Passive : verb agrees with the *karma*
  - Some other cases

raama ko jaanaa paDaa

‘I+to’ ‘go’ ‘had to’

“I had to go”

raama ko calanaa caahiye

‘Ram’ ‘to’ ‘walk’ ‘should’

“I should leave”

Example

Raama jaataa hai

‘Ram’ ‘go+hab’ ‘pres’
“Ram goes”

jaa

karta

raama

- *TAM + Postposition together help in identifying the relations*

Raama ko jaanaa paDaa

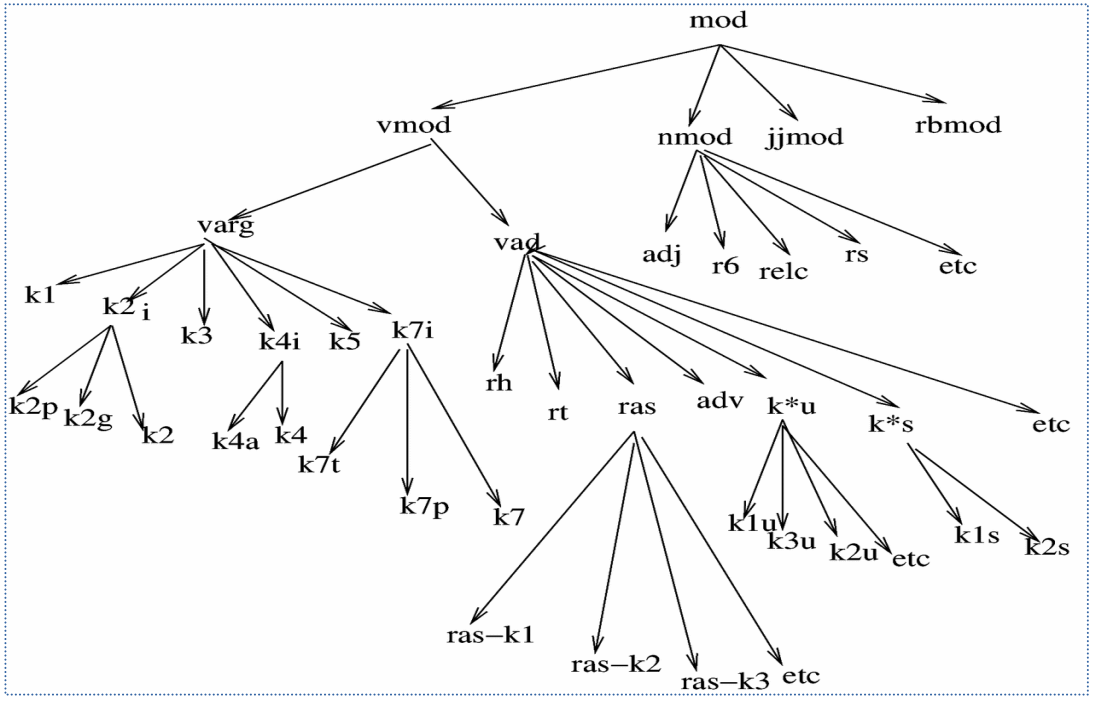
‘Ram+dat’ ‘to go’ ‘had’
“Ram had to go”

jaa

karta

raama

Dependency Relations



Interchunk relations

Parsing Indian languages

- Indian languages
  - Rich morphology
  - Relatively flexible word order

  For example,

  1. a) *baccaa phala khaataa hai*

  ‘child’ ‘fruit’ ‘eat+hab’ ‘pres’

  b) *phala baccaa khaataa hai*

  c) *phala khaataa hai baccaa*

  d) *baccaa khaataa hai phala*

Levels of Analysis

L1 – Semantic relations : karakas, eg *raama* karta

L2 – Morphosyntactic : vibhakti, eg *raama* prathamaa

L3 – Morphological representation (abstract) : vibhakti markers, eg raama + su (Sanskrit)

raama + 0 (Hindi)

raama + du (Telugu)

L4 – Phonological form : raamaH (Sans)

raama (Hindi)

raamudu (Telugu)

Our Model

- Morph analysis
- POS tagging
- Identify minimal constituents (chunks/bags) and their heads
- Mark the relations across chunks (head to head relation)
- Bag-internal dependencies are left unspecified
- The trees can be fully expanded if required

For Example

meraa baDzaa bhaaii bahuta phala
khaataa hai

=>
meraa_PRP baDzaa_JJ bhaaii_NN
bahuta_QF phala_NN khaataa_VM hai_VAUX

=>
((meraa_PRP baDzaa_JJ ***bhaaii****_NN))_NP*
((bahuta_QF ***phala****_NN))_NP*
*((****khaataa****_VM hai_VAUX))_VG*

Example Contd…

*((meraa_PRP baDzaa_JJ **bhaaii**_NN))_NP*
*((bahuta_QF **phala**_NN))_NP*
*((**khaataa**_VM hai_VAUX))_VG*

(t1) *khaa* *(t2)* *khaa*

bhaaii *phala* *bhaaii* *phala*

meraa baDZaa bahuta

The same features can be used for Machine learning as well

Words	***Features***
Sabina	
ne	
is	
chaabii	
se	
taalaa	
khol aa	

Words	***Features***
Sabina	
is	
chaabii	
se	
taalaa	
khol	
rahii_	
hai	

Summary : Some Concepts

- Root and suffix form a word (Prakriti and pratyaya)
- Speaker's intention (vivakshaa)
- Participants in the action denoted by a verbal root (karaka)
- Nominal and verbal inflections (vibhakti)
- Verb-argument agreement (abhihita/ukta)
- Expectancy (aakaankshaa)
- Eligibility (yogyataa)
- Proximity (sannidhi)

Must Read

1. https://cdn.iiit.ac.in/cdn/ltrc.iiit.ac.in/downloads/nlpbook/nlp-panini.pdf _x005F_x005F_x005F_x005F_x005F_x005F_x005F_x0001__x005F_x005F_x005F_x005F_x005F_x005F_x005F_x0001_
2. AnnCorra : Annotation Guidelines for Indian Language Treebanks (posted on Course page on Moodle)

Reference Reading

1. https://www.researchgate.net/publication/228391518_An_introduction_to_dependency_grammar
2. http://www.sfs.uni-tuebingen.de/~dm/10/ss/dep/dg-slides-2x2.pdf
3. https://www.aclweb.org/anthology/W04-2708.pdf
4. https://cl.lingfil.uu.se/~nivre/docs/05133.pdf
5. https://web.stanford.edu/~jurafsky/slp3/15.pdf