A classification of dependency labels in the Hindi Treebank

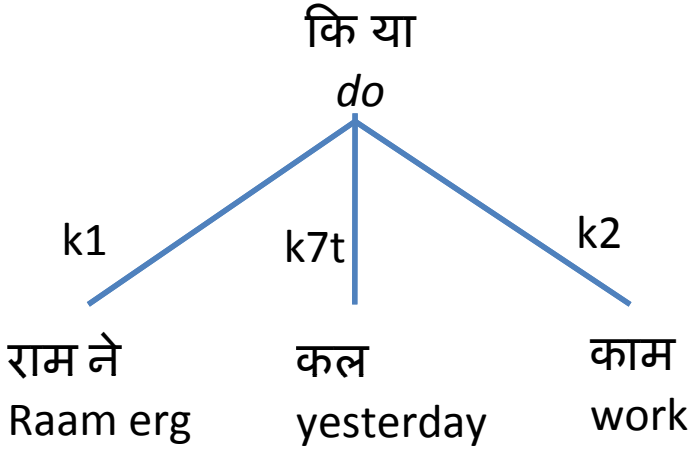
Ashwini Vaidya

Samar Husain

Hindi/Urdu Treebank

- The Hindi/Urdu Treebank is annotated using a dependency grammar framework
- Framework used is CPG: (Computational Paninian Grammar)
  - Panini's 'karaka' theory adapted for annotation scheme
  - A Karaka is the role played by a participant in an action
  - The action in a sentence is denoted through a verb

Dependency labels



Example of a dependency tree. Labels denote relations between a modifier and a modified

- Karakas are the relations between head and child nodes in the treebank
- Relations are depicted between word chunks and not individual tokens
  - E.g. a verb chunk can consist of a finite verb along with its auxiliaries

Examples of karaka labels

- **k1 'karta'** : most independent participant in an event

Rama$_{k1}$ baithaa hai

Rama sit be.pres

'Ram is sitting'

- **k2 'karma'** : locus of the result implied by the verb root

Ram ne **Ravi$_{k2}$** ko dekhaa

Ram erg Ravi acc see.perf

'Ram saw Ravi'

Examples contd.

- **k4 'sampradaana'** : recipient

Ram ne **Mohan**$_{\mathbf{k4}}$ ko kitaab dii

Ram erg Mohan dat book give.perf

'Ram gave Mohan a book

- **k7p 'deshadhikarana'** : location in space

mezja$_{\mathbf{k7p}}$ para kitaab hai

table on book be.pres

'The book is on the table'

Selected dependency labels [TOTAL = 43]

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k7t	temporal location
k7p	spatial location
k1s	noun complement
k2p	destination
pk1, mk1, jk1	Causer, mediator-causer, causee
rh	cause
rt	purpose
rsp	duration
adv	adverb (manner)
pof	Part-of (complex predicates)
ccof	conjunction
fragof	fragment-of

Classification of the tagset

- Our main aim is a classification of the labels in this tagset
- This is desirable for two main reasons:
  - Linguistic : to enable better interpretation of the labels used in the treebank
  - Practical : to be able to generalize these labels to other annotation schemes & facilitate mapping

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k7t	temporal location
k7p	spatial location
k1s	noun complement
k2p	destination
pk1, mk1, jk1	Causer, mediator-causer, causee
rh	cause
rt	purpose
rsp	duration
nmod	noun modification
pof	Part-of (complex predicates)
ccof	conjunction
fragof	fragment-of

Bharti et al (2009) describe a classification scheme for these labels

rh	cause
rt	purpose
rsp	duration
nmod	Noun modification
pof	Part-of (complex predicates)
ccof	conjunction
fragof	fragment-of

The first is the class of all k-labels; includes k1-k7*

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k7t	temporal location
k7p	spatial location
k1s	noun complement
k2p	destination
pk1, mk1, jk1	Causer, mediator-causer, causee
nmod	Nmod (noun modification
pof	Part-of (complex predicates)
ccof	conjunction
fragof	fragment-of

Second is the class of r-labels, including rh, rt, rsp, rd etc.

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k7t	temporal location
k7p	spatial location
k1s	noun complement
k2p	destination
pk1, mk1, jk1	Causer, mediator-causer, causee
rh	cause
rt	purpose
rsp	duration
pof	Part-of (complex predicates)
ccof	conjunction
fragof	fragment-of

- Third, the class of modifier labels: includes nmod, vmod, jjmod

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k7t	temporal location
k7p	spatial location
k1s	noun complement
k2p	destination
pk1, mk1, jk1	Causer, mediator-causer, causee
rh	cause
rt	purpose
rsp	duration
nmod	noun modification

- Finally, the class of non-dependencies consisting of pof, fragof, ccof

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k7t	temporal location
k7p	spatial location
k1s	noun complement
k2p	destination
pk1, mk1, jk1	Causer, mediator-causer, causee
rh	cause
rt	purpose
rsp	duration
nmod	noun modification
pof	Part-of (complex predicates)
ccof	conjunction
fragof	fragment-of

- We get a 4-way classification:
-k-labels
-r-labels
-modifier labels
-other (non dependencies)

Limitations

- There are several limitations to this classification
  - The k-label and r-label distinction fails to capture certain generalizations
  - There is ambiguity in the current annotation practices as a result of the classification

Proposed classification

- Revision of the k-labels and r-labels
- Fine grained distinctions
- Classes based on the label's affinity with the verb (Rambow et al, 2003)
  - Experiments in mapping labels across different frameworks (VerbNet, PropBank, Prague Dependency Treebank's TR)
  - For these purposes, a classification of labels across corpora is defined

Proposed classification

- Compared to Bharati et al (2009)

k1	k2	k3	k4	k5	k2p	k1s	k7p	k7t	rh	rt	nmod	pof	fragof	ccof
K-labels									R-labels		Mod labels	Other (non dependencies)		

- We have a six- way classification

k1	k2	k3	k4	k5	k2p	k1s	k7p	k7t	rh	rt	nmod	pof	frago f	ccof
Invariant Syntactic labels		Local semantic labels					Global semantic labels				Mod labels	‘pof’ type labels		‘ccof’ labels

Invariant syntactic labels

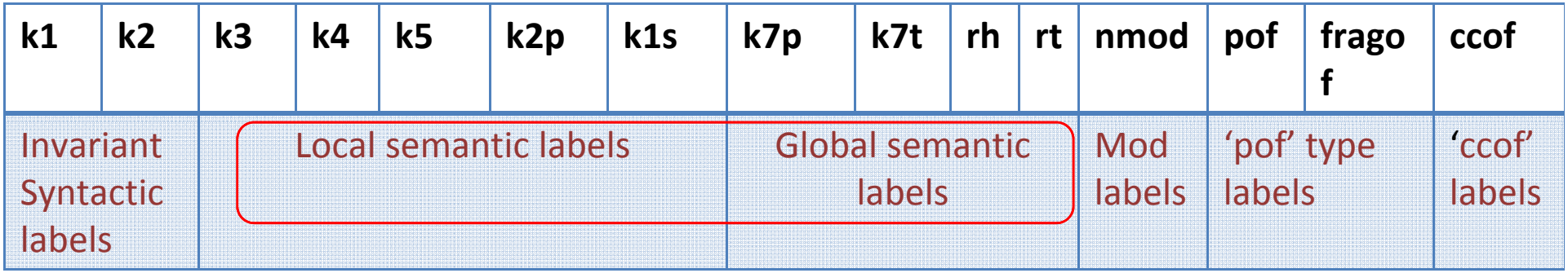
- k1 ‘karta’ and k2 ‘karma’
- Invariant across syntactic alternations like voice
- E.g. aaj khuub mithai khaai gai

  Today many sweets eaten go.pst

  ‘Today many sweets were eaten’
- The label for ‘sweets’ is k2, although ‘sweets’ is now the passivized subject
- This property allows for mapping with PropBank roles Arg0, Arg1

Proposed classification

- Focus on local and global semantic labels



Local semantic labels

- Relation between verb and dependent is ‘local’
- These labels are “relevant to the verb meaning in question”

E.g. Ram ne Mohan ko kahaani sunaai
Ram erg Mohan acc story told.perf
‘Ram told Mohan a story’

- Mohan is a ‘k4’, a beneficiary is a local semantic label
- However, it is a label specific to certain verbs only, e.g. *denaa* ‘to give’, *kahnaa* ‘to say’

Local semantic labels

- Other local semantic labels include
  - k4a ‘anubhava karta’, experiencer, with verbs like *mila* ‘find’, *dikha* ‘see’, *laga* ‘feel’
  - k2p ‘goal’ with verbs like *pahuMca* ‘reach’, *jaa* ‘go’
- The interpretation of these labels is closely bound with the meaning of the verb

Global semantic labels

- These labels are relevant “across different verbs and verb meanings”

E.g maine Mohan ke-liye pustak khariidi

I-erg Mohan for book bought

‘I bought a book for Mohan’

- Here ‘Mohan (for)’ has the label rt or ‘purpose’ which does not change across different verb meanings

Global semantic labels

- More examples of global semantic labels are:
  - k7t ‘location’
  - rh ‘reason’
  - rsp ‘duration’
- Not tied to the meaning of a verb
- A simple distinction between r-labels and k-labels would not be adequate here

Effect on annotation practice

- The distinction between local and global semantic labels will aid in more consistent annotation
- Surface cues can (mis)guide the annotation process. E.g. ‘ko’ for k2, ‘se’ for k3
- An understanding of local and global distinctions will result in less ambiguity in the labels

Local vs. global for k5

- k5 ‘source’ has a broad interpretation

1. Kishori haridwaar se$_{k5(l)}$ dilli aayi thi
   Kishori Haridwar from Delhi come.perf be.pst
   ‘Kishori came to Delhi from Haridwar’

2. UPA sarkar ne vritta mantralaya se$_{k5(g)}$ 30 crore rupayon ki sahaaytaa di
   UPA govt erg upper house from 30 cr rupees gen help give
   ‘The UPA government gave 30 crore rupees from the Upper House as aid’

Local vs. global for k3

- Similar is the case for k3 ‘instrument’

1. Ve pyaaz ko daant se$_{k3(l)}$ kaat-kaat kar khaana pasand karte hai
   he onion acc teeth inst cut-redup-part eat like do be
   ‘He likes to eat the onion by tearing it with his teeth’

2. Jawaan helicopter ke jariye$_{k3(g)}$ logon ki talaash kar rahe the
   soldiers helicopter of means people gen search do prog be
   ‘The soldiers were searching for people with helicopters’

Re-analysis of these labels

- Annotation practices should restrict k3 and k5 to more local semantic usages
- For the global usages, a different label can be used

Summary

- Local vs. global distinctions would identify the core participants in the verb's event
- Labels like k3 and k5 would not be over-loaded
- Mapping to other frameworks that make distinctions between the core and non-core participants will be easier

Extra

Modifier labels

- Labels used to express different kinds of modification. Not always fully specified
- For example, nmod (noun modification) has these subtypes
  - nmod_relc (relative clause)
  - nmod_k1inv (participial relative)
  - nmod (other kinds of noun modification)
- We do not make changes to this particular class of labels

‘pof’ and ‘ccof’

- Certain non-dependencies are tagged with labels like ‘pof’ i.e. part-of or ‘fragof’ fragment-of
- We create a new class of ‘pof’ type labels that includes fragof
- Also, we propose a second class of ‘ccof’ type labels to cover cases of co-ordination and subordination

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k2p	destination
k1s	noun complement
k7p	spatial location
k7t	temporal location
rh	cause
rt	purpose
rsp	duration
nmod	noun modification
pof	Part-of (complex predicates)
fragof	Fragment-of
ccof	conjunction

Invariant syntactic relations

Local semantic relations

Global semantic relations

Modifier relations

pof-type relations

ccof-type relations

k1	karta (similar to agent/doer)
k2	karma (similar to patient/theme)
k3	instrument
k4	beneficiary
k5	source
k2p	destination
k1s	noun complement
k7p	spatial location
k7t	temporal location
rh	cause
rt	purpose
rsp	duration
nmod	noun modification
pof	Part-of (complex predicates)
fragof	Fragment-of
ccof	conjunction

Invariant syntactic relations

Local semantic relations

Global semantic relations

Modifier relations

pof-type relations

ccof-type relations