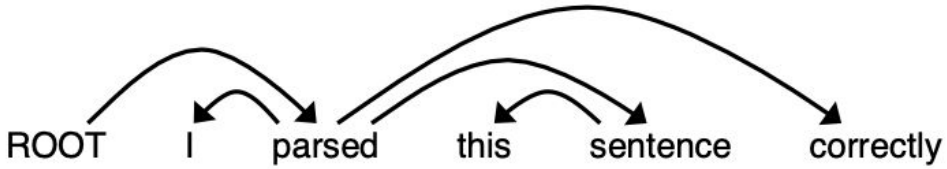
Dependency Parsing - Hindi

Ayan Biswas - 2019121009
Veeral Agarwal - 2019114009
Mentor - Saujas Vaduguru
Course instructor - Prof. Manish Shrivastava

Objective:

- Dependency parsing in Hindi.
- Transition/shift based.
- Use of SVMs for the greedy choice.

Transition Based Dependency Parsing



Stack	Buffer	New dependency	Transition
[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration
[ROOT, I]	[parsed, this, sentence, correctly]		SHIFT
[ROOT, I, parsed]	[this, sentence, correctly]		SHIFT
[ROOT, parsed]	[this, sentence, correctly]	parsed→I	LEFT-ARC

Dataset Used:

We have used the data corpus taken from LTRC, IIIT Hyderabad. The data that was provided was already a parsed data by hindi shallow parser in Shakti Standard Form(SSF). This parsed data was used for training , testing and development.

Preprocessing:

- Merging files and simplification
- Head and Dependency extraction
- Filter non parsable sentences
- Add unrelated dependencies

Initial data

1 ((NP <fs name='NP' drel='nmod:NP2'>

1.1 यहाँ PRP <fs af='यहाँ,pn,,,,o,,' name='यहाँ' posn='10'>

1.2 से PSP <fs af='से,psp,,,,,,' name='से' posn='20'>

))

...

Simplified

H NP NP nmod NP2

T यहाँ PRP यहाँ

T से PSP से

Head Extracted

H यहाँ यहाँ NP PRP NP nmod NP2

H दूरी दूरी NP NN NP2 jjmod JJP

Dependency Extracted

H यहाँ यहाँ NP PRP NP nmod NP2 ; H दूरी दूरी NP NN NP2 jjmod JJP ; R ; nmod

Models: Features Used

All combinations (3C1 + 3C2 + 3C3) = 7 of -

- Head Word - Chunk 1, Head Word - Chunk 2
- POS Tag of Head Words
- Chunk Tags

in the form of One Hot Vectors.

Models: What all are predicted?

- Choice to make, given chunks - Left, Right, Unrelated/Unknown
- Edge Relation Type - nmod, jjmod etc

Models (Cont'd)

- Linear SVM Classifiers are used.
- Total 14 models (7 for each label set)

Features	**L/R/U**	**Edge Relation**
Word	0.77	0.75
POS	0.72	0.61
Chunk	0.77	0.63
Chunk + POS	0.76	0.64
Word + POS	0.77	0.77
Word+ Chunk	0.78	0.77
Word + Chunk + POS	0.78	0.77

Observations:

- Why is 77% so consistent across models?
- POS is possibly the least useful feature.
- Is accuracy a good quantifier?

Classification Details:

Head Word Only

	precision	recall	f1-score	support
L	0.75	0.71	0.73	4174
R	0.66	0.77	0.71	10448
U	0.83	0.78	0.80	21561
accuracy			0.77	36183
macro avg	0.75	0.75	0.75	36183
weighted avg	0.77	0.77	0.77	36183

Chunk Only ->

	precision	recall	f1-score	support
L	0.71	0.83	0.76	4174
R	0.67	0.72	0.69	10448
U	0.84	0.78	0.81	21561
accuracy			0.77	36183
macro avg	0.74	0.78	0.75	36183
weighted avg	0.77	0.77	0.77	36183

POS Only ->

	precision	recall	f1-score	support
L	0.60	0.31	0.41	4174
R	0.67	0.69	0.68	10448
U	0.75	0.81	0.78	21561
accuracy			0.72	36183
macro avg	0.68	0.61	0.63	36183
weighted avg	0.71	0.72	0.71	36183

Results:

Word + POS + Chunk - L/R/U

	precision	recall	f1-score	support
L	0.74	0.79	0.77	4174
R	0.66	0.79	0.72	10448
U	0.86	0.77	0.81	21561
accuracy			0.78	36183
macro avg	0.75	0.78	0.77	36183
weighted avg	0.79	0.78	0.78	36183

Word + POS + Chunk - Edge Relationship Labels

accuracy			0.77	14622
macro avg	0.41	0.37	0.38	14622
weighted avg	0.76	0.77	0.76	14622

Model - Edge Relation	**Macro Average**
Word	0.35
Chunk	0.20
POS	0.21

Observations: Edge Relation

- Some classes have 1100~1500 instances - ROOT, ccof, k2 etc
- Quite a few have < 10 samples - ras-k7p, rbmod etc
- Huge variance in class-wise performance.

Inferences:

- Accuracy is a dangerous metric.
- POS does not seem to contribute useful information the model.
- L/R/U seems to be reliable.
- Edge Relations : Hard to judge, lack indicators to prove reliability

Discussion:

- The features represent ordered pair of words in high dimensional space.
- Dependance is a relation between two words in the sentence.
- In a One Vs All classification mode, it seems that these dependencies are sufficiently far apart to be linearly separated.
- Is this learning any semantic concept, or fitting to a language well by observing and memorizing a large part of it?

Future Work:

- Generate embeddings for words, POS, Chunk tags or entire chunks together.
- Similarity needs to be defined.
- Study similarity - A distance metric vs classification.