Leveraging Newswire Treebanks for Parsing Conversational Data with Argument Scrambling

Riyaz Ahmad Bhat, Irshad Ahmad Bhat, Dipti Misra Sharma

International Institute of Information Technology, India

Table of contents

Introduction

Formal vs. Informal Language Use i

- Formal Language
  - ◇ adherence to defining typological properties of a language
    1. the order of subject, object and verb,
    2. the order of possessive (genitive) and head noun,
    3. the order of adposition and noun, and
    4. the order of adjective and noun.
  - ◇ examples: newswire, academic writing, religious sermons etc.
- Informal Language
  - ◇ less adherence to standard grammar
    1. variation in word-order
    2. implicit information such as ellipsis, null co-ordination etc.
    3. argument scrambling (morphologically rich languages)
  - ◇ examples: colloquial language, conversations, social media, movie scripts etc.

Formal vs. Informal Language Use ii

- **Current NLP is mostly build on Newswire texts:**
  - ◇ newswire texts greatly adhere to standard grammar
  - ◇ more passive sentences, lesser imperative and interrogative sentences

Formal vs. Informal Language Use iii

	News	Fiction	Non Fict.	Blog	Bible	Legal	Medical	Social	Spoken	Wiki	Web	Reviews
Anc. Greek		✓	✓		✓							
Arabic	✓											
Basque	✓	✓										
Bulgarian	✓	✓				✓						
Catalan	✓											
Chinese										✓		
Croatian	✓									✓		
Czech	✓		✓			✓	✓					✓
Danish	✓	✓	✓						✓			
Dutch	✓						✓			✓		
English		✓	✓	✓				✓	✓		✓	✓
Estonian	✓	✓										
Finnish	✓	✓		✓		✓				✓		
French	✓	✓								✓		✓
Galician	✓		✓			✓	✓					
German	✓									✓		✓
Gothic					✓							
Greek	✓								✓	✓		
Hebrew	✓											
Hindi	✓											
Hungarian	✓											
Indonesian	✓			✓								
Irish	✓	✓				✓					✓	
Italian	✓					✓				✓		
Kazakh		✓								✓		
Latin		✓	✓		✓							
Latvian	✓											
Norwegian	✓		✓	✓								
O.Slavonic					✓							
Persian	✓	✓	✓			✓	✓	✓	✓			
Polish	✓	✓	✓									
Portuguese	✓			✓								
Romanian	✓	✓	✓			✓	✓			✓		
Russian	✓	✓	✓							✓		
Slovenian	✓	✓	✓						✓			
Spanish	✓			✓						✓		✓
Swedish	✓	✓	✓						✓			
Tamil	✓											
Turkish	✓		✓									

Formal vs. Informal Language Use iv

Challenges in Parsing Informal Texts using Newswire Annotations

- **Sampling Bias:** limited-sized newswire treebanks only represent a subset of possible structures

S.No.	Order	Percentage
1	S O V	91.83
2	O S V	7.80
3	O V S	0.19
4	S V O	0.19
5	V O S	0.0
6	V S O	0.0

Our Goal

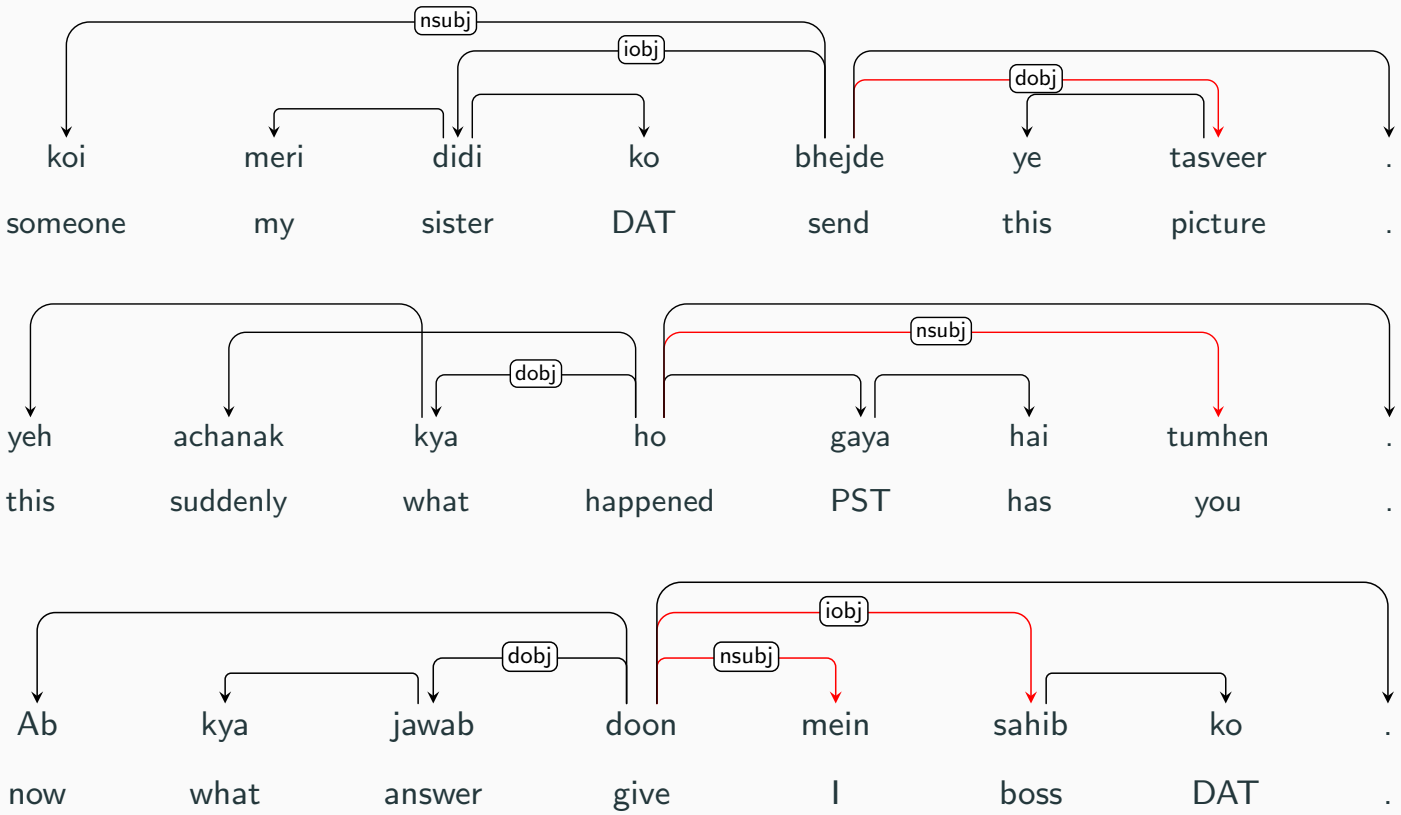
- to fix the sampling bias
- to exploit existing newswire annotations for parsing informal texts

Argument Scrambling

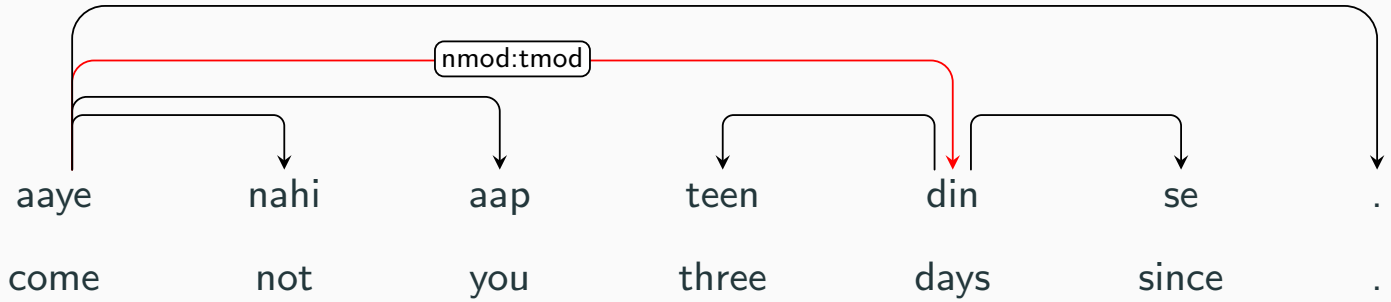
Argument Scrambling i

- movement of arguments for pragmatic reasons (change in information structure)
- default or unmarked constituent order e.g. Hindi $\rightarrow$ SOV
- common in day-to-day conversation
- word-order flexibility exploited for information structure
- free word-order languages allow n factorial (n!) permutations
  (where n is the number of verb arguments and/or adjuncts)

Argument Scrambling ii



Argument Scrambling iii

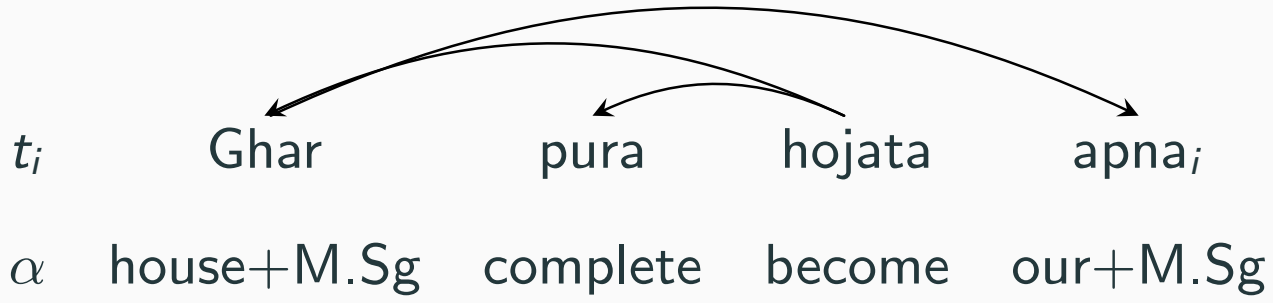


Relation	Newswire (Head Final)	Conversation (Head Final)
Subject	∼100	67
Direct Object	75	64
Indirect Object	100	52
Time	100	58
Place	∼100	53

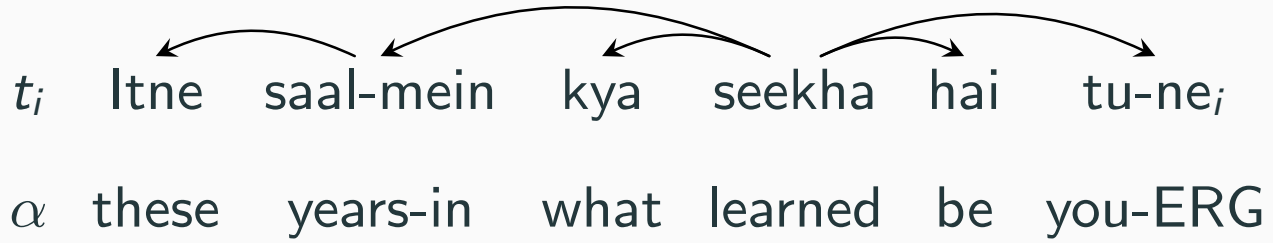
Argument Scrambling iv

What triggers scrambling?

- **Agreement**



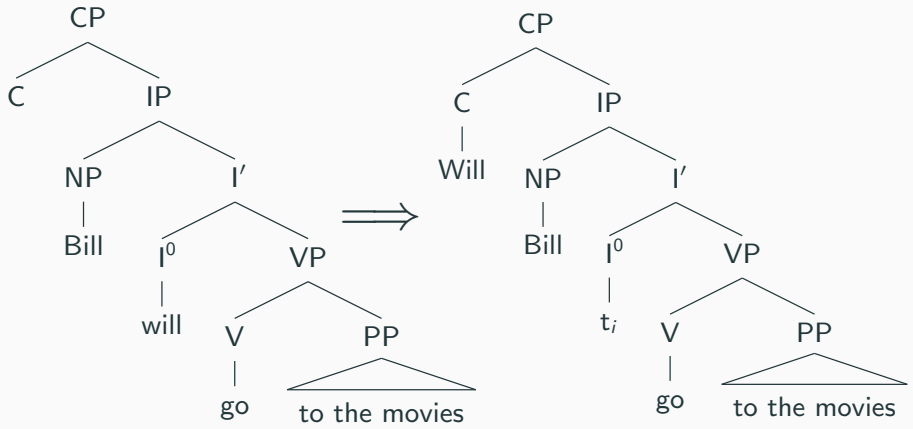
- **Case Marking**



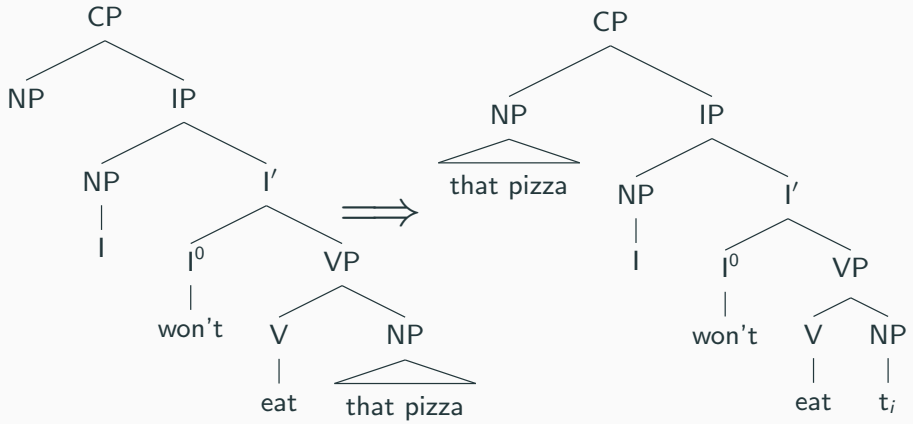
Tree Transformations

Tree Transformations i

- **Transformational grammar:** syntactic variations derivable from basic structures called kernels Chomsky [1]
  - ◇ statements to questions e.g. subject-auxiliary inversion



  - ◇ topicalization

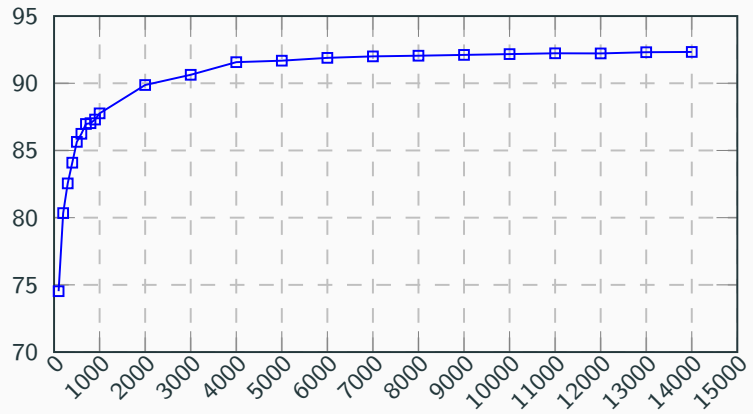


Tree Transformations ii

- **Dependency Tree Transformation:**
  - ◇ permute all the nodes of verbal projections in gold dependency trees
  - ◇ preserve non-projective arcs by pseudo-projectivizing a tree before transformations
  - ◇ only alter the linear precedence relations between arguments of a verbal projection
  - ◇ t x 10! (more than 3 million) possible permutations for 't' syntactic trees containing an average of 10 nodes

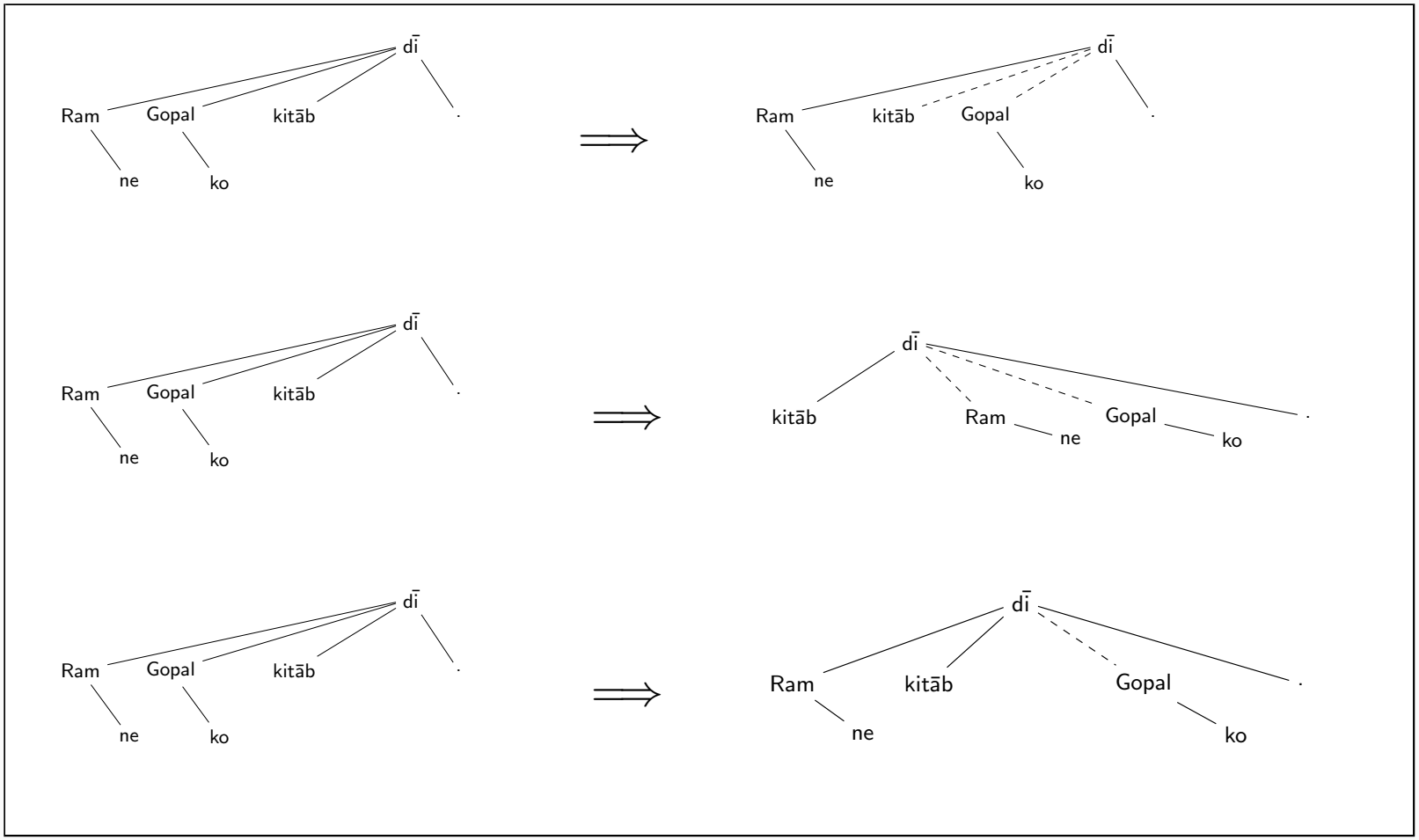
Tree Transformations iii

- ◇ restrict permutations
  - ○ permute a subset of the training data (representative of the newswire domain)



  - ○ take only k permutations with lowest perplexity assigned by a language model
    - ⋆ k is the number of nodes permuted for each projection
    - ⋆ language model is trained on a large and diverse data set (newswire, entertainment, social media, stories, etc.)

Tree Transformations iv



Data

Data i

Newswire Training and Evaluation Data

Training	Testing	Development
13304	1684	1659

Newswire Transformed(Scrambled) Data

- generated 9K trees from 4K representative sentences

Movie Scripts and Twitter Data

- Bollywood movie scripts
- Twitter posts of Hindi monolingual speakers
  - ◇ select Hindi only tweets using a language identification system
  - ◇ only those tweets that contain a minimum of one argument scrambling
  - ◇ CNN-based sentence classification using the canonical and transformed treebank data (∼98%), Yoon Kim [2]

Training	Testing	Development
-	250	250

S.No.	Order	Percentage
1	S O V	33.07
2	O S V	23.62
3	O V S	17.32
4	S V O	14.17
5	V O S	9.45
6	V S O	2.36

Parsing Framework

Transition Systems i

- **Arc-eager algorithm:**
  - ◇ defines $2n$ configurations for $w_1,\ldots,w_n$; $C = (S, B, A)$; where S is a stack, B a buffer, and A a set of arcs
  - ◇ **Initialization:** $S = [\text{ROOT}]$, $B = [w_1,\ldots,w_n]$ and $A = \emptyset$
  - ◇ **Termination:** the buffer is empty and the stack contains ROOT

  Types of transitions (t):

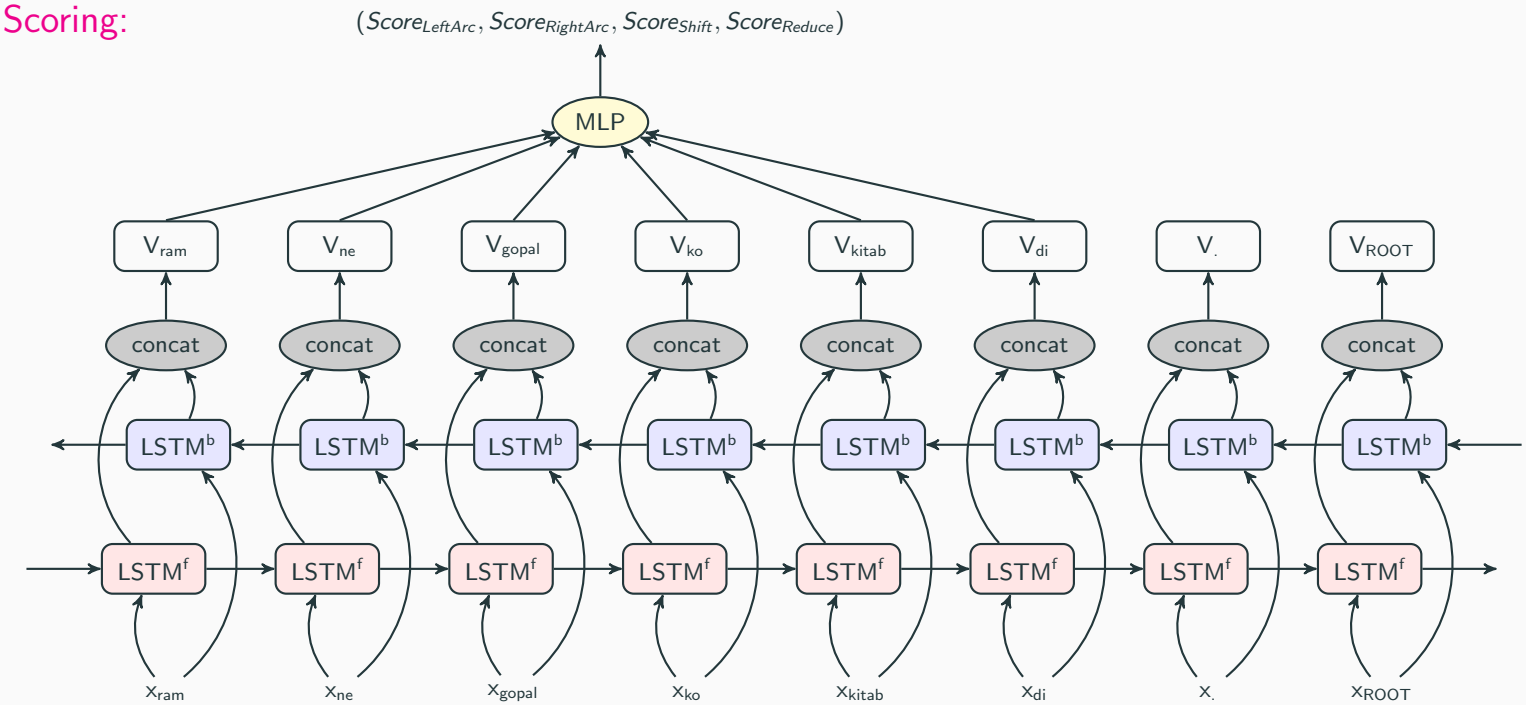
  1. A LEFT-ARC(l) adds an arc $B_j \xrightarrow{l} S_i$ to A & removes S_i from the stack
     - ○ **Precondition:** S_i does not already have a head
  2. A RIGHT-ARC(l) adds an arc $S_i \xrightarrow{l} B_j$ to A & pushes B_j onto the stack
  3. The REDUCE transition removes the top node in the stack
     - ○ **Precondition:** S_0 has a head
  4. The SHIFT transition moves B_0 from buffer to stack

Transition Systems ii

Configuration:

s2	s1	s0	b0	b1	b2
ram	gopal	kitab	di	.	ROOT
↘ ne	↘ ko				

Scoring:



Experiments and Results

Experiments and Results i

- **Training procedure**:
  - ◇ Data Augmentation: train parsing model on the union of multiple views of the treebank
  - ◇ Crosslingual Model: train parsing model on a union of crosslingual treebanks (Hindi+English)

Experiments and Results ii

Data-set	**Newswire**$^{PG\,/\,UD}$		**Newswire**$^{PG\,/\,UD}$**+Transformed Newswire**$^{PG\,/\,UD}$		**Newswire**^{UD}**+English**^{UD}	
	UAS	LAS	UAS	LAS	UAS	LAS
Newswire^{PG}	96.41	92.08	$96.07^{-0.34}$	$91.75^{-0.33}$	-	-
Conversation^{PG}	74.03	64.30	$84.68^{+10.65}$	$73.94^{+9.64}$	-	-
Newswire^{UD}	95.04	92.65	$94.59^{-0.45}$	$92.03^{-0.62}$	$94.56^{-0.48}$	$91.87^{-0.78}$
Conversation^{UD}	73.23	64.77	$83.97^{+10.74}$	$74.61^{+9.84}$	$77.73^{+4.5}$	$68.12^{+3.35}$

Table 1: Parsing results with gold POS-tag

Data-set	**Newswire**$^{PG\,/\,UD}$		**Newswire**$^{PG\,/\,UD}$**+Transformed Newswire**$^{PG\,/\,UD}$		**Newswire**^{UD}**+English**^{UD}	
	UAS	LAS	UAS	LAS	UAS	LAS
Newswire^{PG}	94.55	89.51	$94.29^{-0.26}$	$89.28^{-0.23}$	-	-
Conversation^{PG}	69.52	58.91	$79.07^{+9.55}$	$67.41^{+8.5}$	-	-
Newswire^{UD}	93.85	90.59	$93.32^{-0.53}$	$89.98^{-0.61}$	$93.22^{-0.63}$	$89.72^{-0.87}$
Conversation^{UD}	68.81	59.43	$78.38^{+9.57}$	$67.98^{+8.55}$	$71.29^{+2.48}$	$62.46^{+3.03}$

Table 2: Parsing results with auto POS-tag

- Improvement over Canonical: **8.5 LAS**

Thank You

References i


Noam Chomsky.
Aspects of the Theory of Syntax, volume 11.
MIT press, 2014.


Yoon Kim.
Convolutional neural networks for sentence classification.
arXiv preprint arXiv:1408.5882, 2014.