

# CLPL: Computational Morphology

*an extended version of the ICLAP18 workshop by*

**Tafseer Ahmed & Mutee U Rahman**

# Objectives

- defining Morphological Features
- creating Morphological Analyzers/Generators

# Objectives

- **defining Morphological Features**
  - different schemes and resources
  - noun and verb morphology
  - issues of Pakistani languages
- **creating Morphological Analyzers/Generators**
  - Finite State Transducer
  - Machine Learning
  - Unsupervised learning of morphology

# Morphology - introduction

## •denationalization

- (Arabic) کتاب ، کاتب ، مکتوب ، اقوام
- paRH-vaa-t-aa پڑھاتا
- لکھ، لکھو، لکھا، لکھیے، لکھائو، لکھوائو
- Irregular: go , went, گیا , جا

Morphology: Study of word formation

# Morphology - introduction

uygarlaş, tıramadıklarımızdanmış, sizinizcasına

uygar	civilized
+laş	+BEC
+tır	+CAUS
+ama	+NEGABLE
+dık	+PPART
+lar	+PL
+ımız	+P <sub>1</sub> PL
+dan	+ABL
+mış	+PAST
+sınız	+2PL
+casına	+As_if

'(behaving) as if  
you are among  
those whom we  
could not cause  
to become  
civilized'

# Tokenization & Segmentation

- آسان ترین
- آتش پرست
- ...

Lebensversicherungsgesellschaftsangestellter

Life insurance company employee (German)

# Morphological Features

# Part of Speech (PoS) Tagset

- **PoS** – Noun, Pronoun, Adjective, Verb, Adverb, ...
- **PoS Tags** – N, PN, ADJ, V, ADV, PUNC, EMOJI,
- **PoS TagSet** – Set of tags {N, PN, ADJ, V, ADV, CONJ, PUNC, ..}
- How PoS Tags are linked to Morphology?
  - Morphology of every PoS has certain rules
  - (Many) PoS tagsets also include morphological features  
e.g. (لڑکوں)\_N\_Masc\_PL)



# Universal PoS Tagset

Open class words	Closed class words	Other
<u>ADJ</u>	<u>ADP</u>	<u>PUNCT</u>
<u>ADV</u>	<u>AUX</u>	<u>SYM</u>
<u>INTJ</u>	<u>CCONJ</u>	<u>X</u>
<u>NOUN</u>	<u>DET</u>	
<u>PROPN</u>	<u>NUM</u>	
<u>VERB</u>	<u>PART</u>	
	<u>PRON</u>	
	<u>SCONJ</u>	

<https://universaldependencies.org/u/pos/index.html>

# Morphological Features

- ParGram

## Nominal Features

ANIM:  $\pm$

CASE: ACC, DAT, GEN, NOM

GEND: FEM, MASC, NEUT

NTYPE: COUNT, MASS, PROPER

NUM: PL, SG

PERS: 1, 2, 3

REFL:  $\pm$

# Morphological Features

- **Universal Tagset**

Lexical features	Inflectional features	
	<i>Nominal*</i>	<i>Verbal*</i>
<u>PronType</u>	<u>Gender</u>	<u>VerbForm</u>
<u>NumType</u>	<u>Animacy</u>	<u>Mood</u>
<u>Poss</u>	<u>NounClass</u>	<u>Tense</u>
<u>Reflex</u>	<u>Number</u>	<u>Aspect</u>
<u>Foreign</u>	<u>Case</u>	<u>Voice</u>
<u>Abbr</u>	<u>Definite</u>	<u>Evident</u>
	<u>Degree</u>	<u>Polarity</u>
		<u>Person</u>
		<u>Polite</u>
		<u>Clusivity</u>

<https://universaldependencies.org/u/feat/index.html>

# Morphological Features

## •Unimorph

Number	Dual	DU
Number	Greater paucal	GPAUC
Number	Greater plural	GRPL
Number	Inverse	INVN
Number	Paucal	PAUC
Number	Plural	PL
Number	Singular	SG
Gender	Bantu Noun Classes	BANTU1-23
Gender	Feminine	FEM
Gender	Masculine	MASC
Gender	Nakh-Daghestanian Noun Classes	NAKH1-8
Gender	Neuter	NEUT

# Morphological Feature Schemes - Discussion

# Morphology of Nouns and Verbs

# (Urdu) Nouns

Form/Case	Number	Gender	
Nominal	Singular	Masculine	لڑکا آیا
Nominal	Plural	Masculine	دو لڑکے آئے
Oblique	Singular	Masculine	لڑکے نے کہا
Oblique	Plural	Masculine	دو لڑکوں نے کہا

# (Urdu) Nouns

Form/Case	Number	Gender		
Nominal	Singular	Masculine	سیب گرا	لڑکا آیا
Nominal	Plural	Masculine	دو سیب گرے	دو لڑکے آئے
Oblique	Singular	Masculine	سیب کو گرایا	لڑکے نے کہا
Oblique	Plural	Masculine	دو سیبوں کو گرایا	دو لڑکوں نے کہا



# (Urdu) Nouns

Case/Form	Number	Gender		
Nominal	Singular	Feminine	لڑکی آئی	لڑکا آیا
Nominal	Plural	Feminine	دو لڑکیاں آئیں	دو لڑکے آئے
Oblique	Singular	Feminine	لڑکی نے کہا	لڑکے نے کہا
Oblique	Plural	Feminine	دو لڑکیوں نے کہا	دو لڑکوں نے کہا

# (Urdu) Nouns

Case/Form	Number	Gender		
Nominal	Singular	Feminine	عورت آئی	لڑکی آئی
Nominal	Plural	Feminine	دو عورتیں آئیں	دو لڑکیاں آئیں
Oblique	Singular	Feminine	عورت نے کہا	لڑکی نے کہا
Oblique	Plural	Feminine	دو عورتوں نے کہا	دو لڑکیوں نے کہا

# (Urdu) Nouns

## **borrowed plurals and plural Rules:**

- درجے، درجات
- کمپیوٹروں، کمپیوٹرز
- کتابیں، کتب
- ملک، ممالک
- -----

**vocative form, phonological/orthographic rules**

# (Urdu) Verbs

Following are the forms of an English verb:

- write, writes, wrote, written, writing

How many forms does an Urdu Verb e.g. likH (لکھ) have?

# (Urdu) Verbs

How many forms does an Urdu Verb e.g. likH (لکھ) have?

# (Urdu) Verbs

How many forms does an Urdu Verb e.g. likH (لکھ) have?

لکھ، لکھا، لکھی، لکھے، لکھیں، لکھتا، لکھتی، لکھتے،  
ڈھونڈتیں، لکھنا، لکھنی، لکھنے، لکھنیں، لکھوں، لکھیں،  
لکھو، لکھئے

# (Urdu) Verbs

	Form	Person	Gender	Number
لکھ	root			
لکھتا	imperfective		masc	singular
لکھتیں	imperfective		fem	plural
لکھی	perfective		fem	singular
لکھوں	subjunctive	1st		singular
لکھیے	percative	2nd		

# Nouns and Verbs

What are the noun and verb forms and corresponding morphological features in other Pakistani languages?



# Derivational Morphology

# Comp. Morph. of Pakistani languages - resources

# Issues of Pakistani Languages

# Issues related to Pakistani Languages

## **Rich morphology**

- Inflections, suffixations, infixations, .....
- 48+ inflections of verbs in Urdu,
- 75+ inflections of Verbs in Sindhi...

# Issues related to Pakistani Languages

Nominative and Oblique forms with ambiguity

- laRkE+N+Masc+Pl+Nom
- laRkE+N+Masc+Sg+Obl

# Issues related to Pakistani Languages

## Inflectional case

- Ablative Case (او گھروں آیا)
  - With nouns: گھران، اسکو لان، گھروں، سکولوں
  - With Adverbs: اپروں، ہیٹھوں،

**Issue:** Which morphological feature (case or form) should be used for obliques and inflectional cases?

# Issues related to Pakistani Languages

## •Morphological causatives

- khAyA, khilAyA, khilwAyA (Urdu) (کھایا، کھلایا، کھلوایا)
- likhyO, likhAyO, likhArAyO, likhArArAyO, likhArArArAyO... (لکيو، لکاريو، لکارايو، لکارارايو - *Sindhi*)

# Issues related to Pakistani Languages

- Pronominal suffixes - Sindhi
  - putu (son) -> putu<sup>m</sup> (my son),
  - likh<sup>yO</sup> (wrote),
  - likh<sup>yOmANsi</sup> (I wrote to him)



# Issues related to Pakistani Languages

- Reduplication and echo
  - kitaab\_v/shitaab
  - chalte\_chlate
- ???

# PakMorph

a possible collaborative project

# Objectives

- **defining Morphological Features**
  - different schemes and resources
  - noun and verb morphology
  - issues of Pakistani languages
- **creating Morphological Analyzers/Generators**
  - Finite State Transducer
  - Machine Learning
  - Unsupervised learning of morphology