

# Bangla Verb Morphology and a Multilingual Computational Morphology Framework for PC- KIMMO

---

Md. Zahurul Islam & Mumit Khan  
Center for Research on Bangla  
Language Processing (CRBLP)  
BRAC University  
Bangladesh

# Outline

---

- Morphological Analysis
- Bangla Morphology
- Bangla Verb Morphology
- PC-KIMMO
- JKimmo

# Morphological Analysis

---

- Identification of stems from a full word form
- Fundamental requirement for NLP
- Applications
  - Spelling Checker
  - Text to Speech
  - Grammar Checker
  - Machine translation

# Morphological Analysis (cont...)

---

- Primary Components
  - Generative rules
  - Engine
  - User Interface

# Bangla Morphology

---

- Bangla has word boundary and sentence boundary
- Bangla Words are combination of
  - Noun root (নাম প্রকৃতি)
  - Verb root (ক্রিয়া প্রকৃতি)
  - Formative Affixes (প্রত্যয়)
  - Inflections (বিভক্তি)

# Bangla Morphology (cont...)

---

- Formative Affixes (প্রত্যয়)
  - Primary Affixes (কৃত প্রত্যয়)
  - Secondary Affixes (তদ্ধিত প্রত্যয়)
- Inflections (বিভক্তি)
  - Nominal and Pronominal Inflections (শব্দ বিভক্তি)
  - Verbal Inflections (ক্রিয়া বিভক্তি)

Any prefix in Bangla?

# Bangla Morphology (cont...)

---

- Small number prefixes
- Prefixes can be placed before nouns and adjectives
- Bangla Prefixes
  - Only 19 pure Bangla prefixes  
Example: সু (su), সুদিন (good day)
  - Also inherits 26 prefixes from Sanskrit and 13 from other foreign word
  - Total 58 prefixes

Example:

Sanskrit Prefix: উপ (upo), উপহার (Gift)

Farshi Word: বে (bea), বেনামী (anonymous)

# Verb Morphology

---

- All Bangla verbs are generated from verb roots.
- Verb roots are not free morpheme
- √ Sign is used to indicate verb roots
- Verbs are the combination of verb roots, primary affixes and verbal inflections, for example:

করলাম - > √কর্+ল+আম

√কর্ (kor) verb root

ল (lo) Primary Affix

আম (am) Verb inflection



# Verb Morphology (cont...)

---

- By considering similarity, we can categorize verbs into different categories
  - 1500 verb roots
  - 27 categories
- Category Examples:
  - √আ (aa), √হ (ha) etc.
  - √কর (kor), √চল (chal) etc.
  - √আউলা (auula) etc
  - √আরম্ভ (aaromvo)

# Verb Morphology (cont...)

---

## কর (kor) Category

- 28 verb roots
- Characteristic
  - This category is the combination of two letters
  - First letter will be any vowel other than আ (aa) or any consonant without kar symbol (কার- চিহ্ন) or আ (aa). For example: √ উঠ (uot), √ কর (kor), √ কিন্ (kin).
  - Last letter must be consonant with Virama sign.

# Verb Morphology (cont...)

কাল	নাম পুরুষ		মধ্যম			উত্তম
	সাধারণ (সে)	সম্মানাত্মক (তিনি)	সম্মানাত্মক (আপনি)	সাধারণ (তুমি)	বৃহৎ (বুধি)	
বর্তমান						আমি
সাধারণ	করে	করেন	করেন	কর	করিস্	করি
ঘটমান	করছে	করছেন	করছেন	করছ	করছিস	করছি
পুরাযচিত	করেছে	করেছেন	করেছেন	করেছ	করেছিস	করেছি
অনুজ্ঞা	করক	করুন	করুন	কর	কব্	

Table 1: Present form of কর (Kor) Category

# Verb Morphology (cont...)

---

- Verb root + ই (i) = generated verb will be present simple and first person. example:
  - √কর (kor) + ই = করি (I do)
  - √চল (chol) + ই = চলি (I go)
  - √পড় (por) + ই = পড়ি (I read)
- Verb root + ে = generated verb will be present simple and third (familiar) person. For example:
  - √কর (kor) + ে = করে (He does)
  - √চল (chol) + ে = চলে (He goes)
  - √পড় (por) + ে = পড়ে (He reads)

# Verb Morphology (cont...)

---

- Derivation is also available in Bangla.  
Example:
  - কেনা (to buy) -> কেনানো (to make buy)
  - গোয়লা (milk man) -> গোয়ালিনি (milk woman)
  - মামা (maternal uncle), মামি (maternal aunty)
- Compounds are formed by adding of a verb root or an auxiliary to a noun in accusative or locative or an unaffected verbal conjunctive or a verbal participle. Example:
  - পানি দেওয়া (to water)
  - গায়ে-মাখা (to smear on the body)
  - করতে লাগা (to start doing)

# Verb Morphology (cont...)

করলাম ->

কর+ল+াম

কর+সাধারণ+অতীত+প্রথম

করেছিলাম ->

কর+য়েছ+িল+াম

কর+পুরাঘটিত+অতীত+প্রথম

করব ->

কর+ব

কর+সাধারণ+ভবিষ্যৎ+প্রথম

Recognition

কর+ল+াম -> করলাম

কর+য়েছ+িল+াম -> করেছিলাম

কর+ব -> করব

Generation

Lexical form: কর+ল+াম

Surface form: করলাম

# PC-KIMMO

---

- Based on Kimmo Koskenniemi's two level morphology
- Components
  - Two level rules
  - Lexicon
  - Grammar

# PC-KIMMO

---

## □ Limitations

- Allow only Latin script for input and output
- No Graphical User Interface



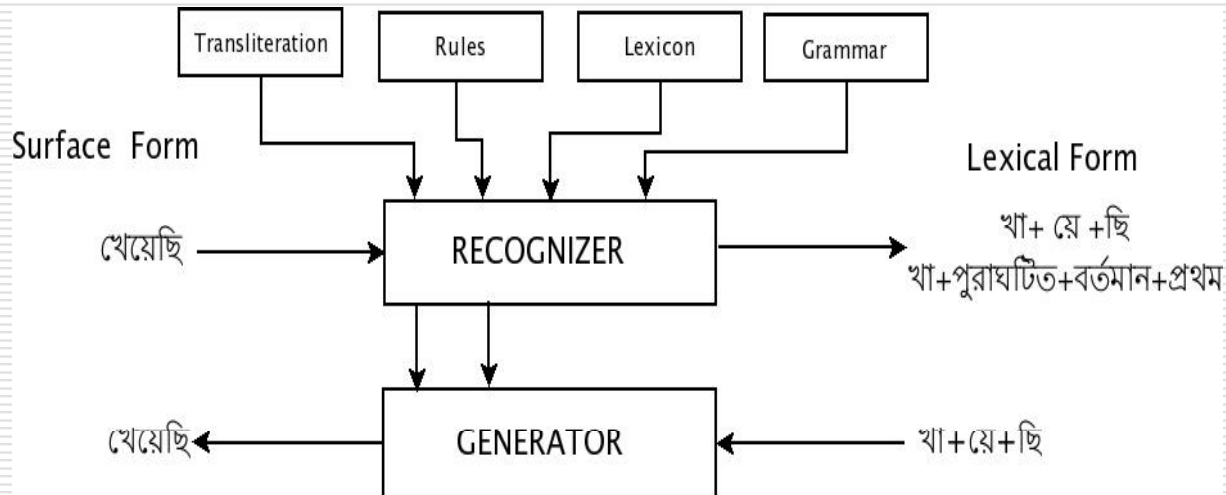
# JKimmo

---

- ❑ A multilingual wrapper around PC-KIMMO
- ❑ Allow us to use Unicode for input and output
- ❑ Localized multilingual Interface

# JKimmo

- Additional component
  - Transliteration File



## JKimmo components

# JKimmo Tutorial

---

## □ Transliteration File

Bangla	Latin	Bangla	Latin	Bangla	Latin	Bangla	Latin	Bangla	Latin
এ	^	আ	a	গ	G	ণ	N	ব	r
অ	A	ি	I	য	G	ভ	t	ল	l
আ	F	ী	I	ঙ	?	থ	T	শ	S
ঊ	H	ূ	u	ঢ	C	দ	d	ষ	\$
ঋ	L	্	U	ছ	C	ধ	D	স	o
ৗ	M	্	R	জ	J	ন	n	হ	h
ঔ	Q	ে	e	ঝ	J	প	p	ড়	'
ঋ	V	ৈ	E	ঞ	Q	ফ	P	ঢ়	"
এ	W	ো	o	ট	V	ব	b	য়	Y
ঐ	X	ৌ	O	ঠ	W	ভ	B	ং	%
ও	Z	ক	k	ড	X	ম	m	ঃ	&
ঔ	F	খ	K	ঢ	Z	য	y	ঁ	~

# JKimmo Tutorial

---

- Rule File
  - KGEN (Rule Compiler)
- Lexicon file
- Grammar
- Interface localization

# Result

The screenshot shows the JKimmo Java Interface of PCKIMMO. The window title is "JKimmo: Java Interface of PCKIMMO". The interface is divided into four main sections:

- ব্যাকরণ (Grammar):** Contains a rule definition for Rule 13: `Person -> DITIOGOU`. The rule body includes: `<Person person> = <DITIOGOU person>`, `<Person aspect> = <DITIOGOU aspect>`, and `<Person time> = <DITIOGOU time>`.
- সূত্র (Table):** A table with 10 rows and 7 columns of numerical data. The rows are labeled 2 through 10. The data is as follows:

2:	0	3	2	3	1	1	3	1
3:	0	4	2	4	1	10	4	1
4:	0	1	2	1	5	1	1	1
5:	8	6	2	1	1	1	1	1
6:	0	1	2	1	1	7	1	1
7:	0	1	2	1	1	1	0	1
8:	0	0	0	0	0	9	0	0
9:	0	0	0	0	0	0	1	0
10:	0	4	2	4	1	1	4	1
- ফলাফল (Results):** Shows the Bengali translation of the rule: `করেহিলাম -> কর+য়েছ+হিল+এম`, `কর+পূর্বস্বত্ব+অতিত+প্রথম`, `খেয়েহিলাম -> খা+য়েছ+হিল+এম`, `খ+পূর্বস্বত্ব+অতিত+প্রথম`, `করব -> কর+ব`, `কর+সাম্বরণ+অবিষত+তৃতীয়`.
- শব্দভান্ডার (Lexicon):** Lists field codes: `FIELDCODE lf U ;lexical item`, `FIELDCODE lx L ;sublexicon`, `FIELDCODE alt A ;alternation`, `FIELDCODE fea F ;features`, `FIELDCODE gl G ;gloss (root)`. It also includes `INCLUDE banglalexicon.lex`.

At the bottom, there is an input field labeled "ইনপুট" with the value "করব", and three buttons: "উৎপাদন করো", "শব্দভুক্ত করো", and "প্রস্থান".

# Summary & Conclusion

---

- ❑ Morphological rules for Bangla Verbs
- ❑ Reusable and robust open-source framework for computational morphology
- ❑ Unicode enabled
- ❑ Multilingual Interface
- ❑ It's free and available at  
([www.bracu.ac.bd/research/crb1p/download.php](http://www.bracu.ac.bd/research/crb1p/download.php))