

Development of Sindhi Lexical Functional Grammar

Mutee U Rahman & Hameedullah Kazi
Isra University, Hyderabad

Outline

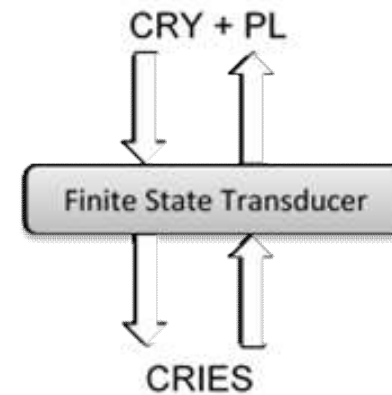
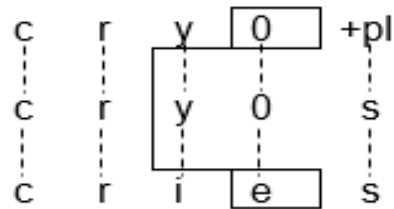
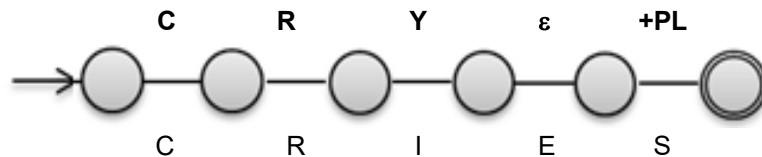
- Introduction
 - Background
 - Finite State Morphology
 - Lexical Functional Grammar
- Overall Development Model
- Implementing Sindhi Morphology
- Implementing Sindhi Syntax
- Coverage
- Conclusion

Background

- Presented work is about development of Sindhi Grammar
- Frameworks used include: Finite State Morphology and Lexical Functional Grammar
- Xerox Finite State Morphology Tools (XFST) and Xerox Linguistic Environment (XLE) are used for Implementation

Finite State Morphology

Singular	Intermediate	Plural	Rule
CRY	CRYS	CRIES	$y \rightarrow ie / \wedge ___ s\#$



Lexical Functional Grammar

- Grammar based on generative grammars (Steedman, 1989), (Dalrymple, 2001)
- Defines linguistic structure at three different levels
 - Lexicon
 - C-structure (Constituent Structure)
 - F-structure (Functional Structure)

Lexical Functional Grammar

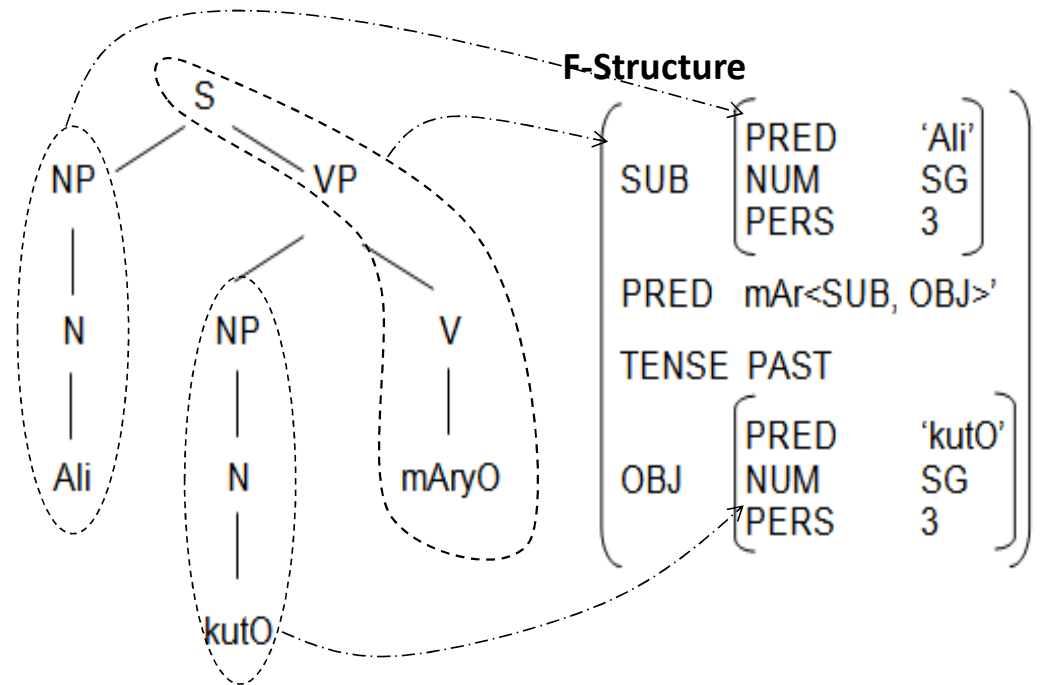
Lexicon

mAryO V (\uparrow PRED) = 'mAru<(\uparrow SUBJ), (\uparrow OBJ)>
 (\uparrow TENSE) = Past
 (\uparrow SUBJ NUM) = SG
 (\uparrow SUBJ PERS) = 3

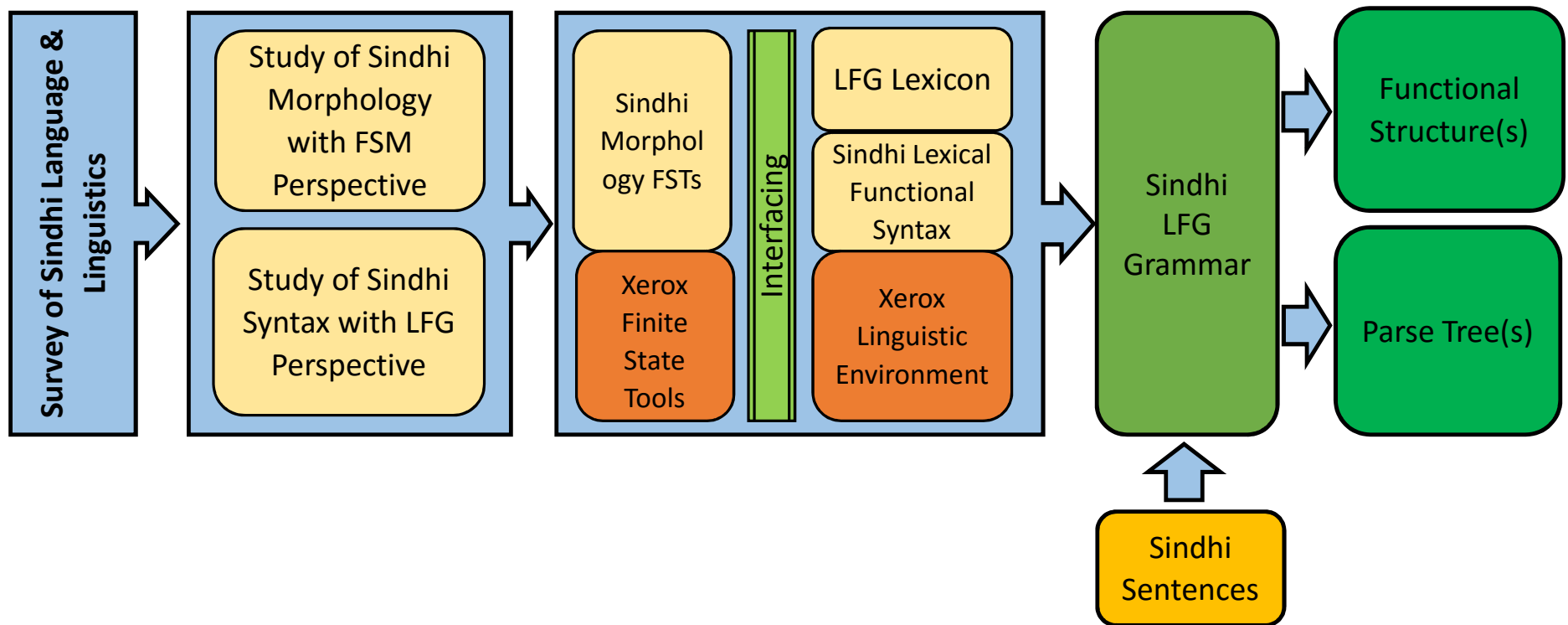
Ali N (\uparrow PRED) = 'Ali'
 (\uparrow NUM) = SG
 (\uparrow PERS) = 3

C-Structure Rules

1. S \rightarrow NP VP
 (\uparrow SUBJ = \downarrow) $\uparrow = \downarrow$
2. NP \rightarrow N
 $\uparrow = \downarrow$
3. VP \rightarrow NP V
4. -----



Grammar Engineering Process



Implementing Morphology

- Morphological paradigms of different POS classes are modeled by incorporating the inflection rules in FSTs using XFST scripts

Implementing Morphology

!SINDHI NOUN MORPHOLOGY

Multichar_Symbols

+Noun +Adjective +Adverb +Verb

+Common +Proper +Abstract !Noun Types

+Animate +Inanimate !Noun Concept

Upper: CHOkir+Noun+Common+Count+Animate+Sg+Masc+Nominative

Intermediate: CHOkir

Lower: CHOkirO

Nouns;

LEXICON Nouns

!Boy (Animate Common Noun)

CHOkir+Noun+Common+Count+Animate; CHOkir N_Cat1;

...

LEXICON N Cat1

+Sg+Masc+Nominative:O #;

+Sg+Masc+Oblique:E #;

+Sg+Masc+Vocative:A #;

+Sg+Fem+Nominative:Ia #;

Morphological analysis of surface form "CHOkirO"

CHOkir {"+Noun" "+Common" "+Count"
"+Animate" "+Sg" "+Masc"
"+Nominative"}

Implementing Morphology

- Following inflections are handled (wherever applicable)
 - Number (CHOkirO, CHOkirA)
 - Gender (CHOkirO, CHOkirIa)
 - Case (CHOkirO, CHOkirE)
 - Tense (likHu, likHAN, likHiyO)
(AhE, huO, hUNdO)
 - Aspect (likHu, likHando)
 - Mood (likHu, likHijANi)
-
- Noun, Pronoun, Ajd, Adv, Postposition, Verb
- Verb

Tense Aspect and Mood not yet analyzed by Sindhi Grammarians

Noun Cases

- Case

Case	Case Marker	Example	
Nominative	∅	CHOkirO	CHOkirO
Accusative / Dative	-E	CHOkirO	CHOkir-E
Postpositional	-E		CHOkir-E
Locative	-E		CHOkir-E
Instrumental	-E		sONT-E sAN
Possessive / Genetive	-E		CHOkir-E JO
Ablative	-AN	gHaru	gHar-AN:
Vocative	-A	CHOkirO	CHOkirA

Oblique Form

- Noun case morphology is further complicated by number and gender inflections in combination with cases



Pronouns

- Pronouns are declined for number and gender
- Marked by **Nominative**, **Oblique** and **Genitive** Cases



Case	Masculine	Feminine
Nom.Sg.	kehRO: CHOkirO	kehRI CHOkirI
Nom.pl.	kehRA CHOkirA	kehRyUN CHOkirUN
Obl.sg	kehRE CHOkirE	kehRIa CHOkirIa
Obl.pl	kehRani CHOkirani	kehRiyuni CHOkiruni
Gen.sg.	muhINjO CHOkirO	muhINJI: CHOkirI
Gen.pl.	muhINjA CHOkirA	muhINjUN CHOkiriUN

Pronominal Suffixes

- Sindhi is one of few Indo-Aryan languages with pronominal suffixes
- Three types of pronominal suffixes are

S.No.	Pronominal Suffix Type		Syntactic Role	Example	
1	Nominal Suffix	اسميه ضمير متصل	Noun	پٽم، پٽس، چاڇهين	puTa-mi
2	Verbal Suffix	فعليه ضمير متصل	Verb	ماريانس، اٿتون، لڪندم	mAri-yAN-si
3	Postpositional Suffix	جري ضمير متصل	Pronoun	کين، ساٿس، وتتون	kHE-na

Verbs

- Verbs are further classified into
 - **Main Verbs** (Transitive & Intransitive)
 - **Compound / Complex Verbs**
 - **Participles** (Present Participle, Past Participle, Future Participle, Verbal Noun, Conjunctive Participle)
 - **Infinitives**
 - **Auxiliary**
 - **Copula**
 - **Modal**

Implementing Syntax

- Nominal Elements
 - Nouns, Pronouns, Adjectives, Adverbs
 - Phrases constituted by above elements
 - Complicated by coordination, postpositional phrases and relative clauses and Cases Marking
- Verbal Elements
 - Verb Subcategorization
 - SUBJ, OBJ, OBJ2, OBL, PREDLINK, COMP, XCOMP
- Adjuncts
 - ADJUNCT, XADJUNCT (Open Adjuncts)

NP Constructions

- Noun (CHOkirO)
- Pronoun-Noun (ihO CHOkirO)
- Adj-Noun (suTHO CHOkirO)
- Pronoun-Ajd-Noun
(ihO suTHO CHOkirO)

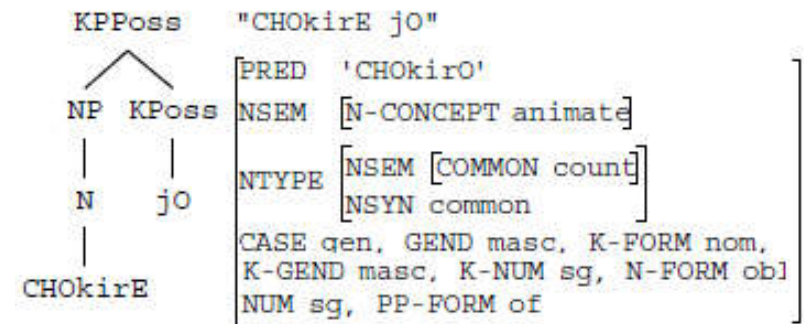
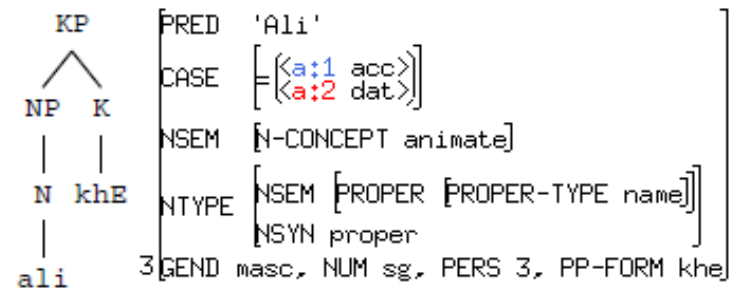
```

"ihO CHOkirO"
  [
    PRED 'CHOkirO'
    NSEM [N-CONCEPT animate]
    NTYPE [
      NSEM [COMMON count]
      NSYN common
    ]
    SPEC [
      DET [
        PRED 'ihO'
        NTYPE [NSYN pronoun]
        CASE nom. DEIXIS proximal.
        GEND masc, NUM sg, PRON-TYPE demon
      ]
    ]
  ]
38
105
107 CASE nom, GEND masc, NUM sg
  
```


Case Marking

- Syntactic Case Marking is handled by using special Case Phrase KP (Bogel., et al, 2009)
- Accusative & Dative Case with “khE” marker
- Genitive case is special as it holds agreement
 - KP Poss is used

F-structure chart
"ali khE"



Verb Subcategorization

SUBJECT & OBJECT

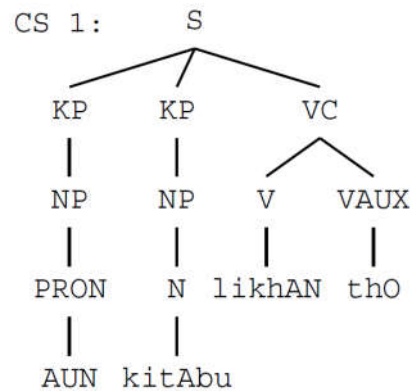
ali kitAbu likHE tHO

Ali.NN book.NC write.Aoirst be.Aux.Pres

Ali Writes a book.

(↑ PRED)='LIKHU<(↑ SUB) (↑ OBJ)>'

"AUN kitAbu likhAN thO"



```

[PRED 'likhu<[1:AUN], [34:kitAbu]>'
SUBJ [
  [PRED 'AUN'
  NTYPE [NSYN pronoun]
  1 [CASE nom, GEND masc, NUM sg, PERS 1, PRON-TYPE personal]
]]
OBJ [
  [PRED 'kitAbu'
  NSEM [N-CONCEPT inanimate]
  NTYPE [
    NSEM [COMMON count]
    NSYN common
  ]
  34 [CASE acc, GEND masc, N-FORM obl, NUM sg]
]]
TNS-ASP [MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]
67 [AUXTYPE tho, GEND masc, NUM sg, VTYPE main]
  
```

Verb Subcategorization

SUBJECT Only

ali **dORE** **tHO**
Ali.NN run.Aoirst be.Aux.Pres
Ali runs.
(↑PRED)='dORi<(↑SUB)'

Verb Subcategorization

Passives: SUBJ → NULL, OBJ → SUBJ

kitAbu likHijE thO

book.NC write.Pass.Aorist be.Aux.Pres

Book is being written/Book writing takes place.

(↑PRED)='LIKHU<(NULL ↑SUB)>'

"kitAbu likhijE"

kitAbu likHibO AhE

book.NC write.Pass.Fut is.Aux.Pres

Book writing takes place.

(↑PRED)='LIKHU<(NULL ↑SUB) >'

```
[PRED 'likhu<NULL, [1:kitAbu]>'
  [PRED 'kitAbu'
   NSEM [N-CONCEPT inanimate]
  ]
SUBJ [NTYPE [NSEM [COMMON count]
             [NSYN common]
          ]
      1[GEND masc, N-FORM obl, NUM sg]
      35[GEND masc, NUM sg, PASSIVE +, VTYPE main]
```

Verb Subcategorization

Passives: NULL Arguments

likhibO

AhE

write.Pass.Fut.Sg.Masc is.Aux.Pres.Sg

Writing takes place.

(↑PRED)='LIKHU<(NULL)>'

likhi jE

tHO

write.Pass.Aorist.Sg be.Aux.Pres.Sg.Masc

(It's) being written.

(↑PRED)='LIKHU<(NULL)>'

Verb Subcategorization

Object-2 (OBJ- θ , Secondary OBJ)

ali CHOkirE=khE KHatu likhE
Ali boy.Obl=dat letter.Nom write

(\uparrow PRED)='likhu<(\uparrow SUB) (\uparrow OBJ2) (\uparrow OBJ)>'

SUB: ali

OBJ2: CHOkirO

OBJ: KHatu

F-structure #1

"ali CHOkirE khE KHatu likhE"

```
PRED 'likhu<[1:Ali], [35:CHOkirO], [71:KHat]>'  
SUBJ [PRED 'Ali'  
NSEM [N-CONCEPT animate]  
NTYPE [NSEM [PROPER [PROPER-TYPE name]]  
NSYN proper  
1 GEND masc, NUM sg, PERS 3  
OBJ2 [PRED 'CHOkirO'  
NSEM [N-CONCEPT animate]  
NTYPE [NSEM [COMMON count]  
NSYN common  
35 CASE dat, GEND masc, N-FORM obl,  
NUM sg, PP-FORM khe  
OBJ [PRED 'KHat'  
NSEM [N-CONCEPT inanimate]  
NTYPE [NSEM [COMMON count]  
NSYN common  
71 CASE nom, GEND masc, NUM sg  
TNS-ASP [TENSE-FORM aorist]  
GEND masc, NUM sg, VTYPE main]
```

Verb Subcategorization

Oblique

tUN CHOkirE=khE ali=khAN KHatu likhArAi
you boy=dat ali=abl letter write.caus2

(↑PRED)='khAu<(↑SUB) (↑OBL) (↑OBJ2) (↑OBJ)>'

SUB: tUN

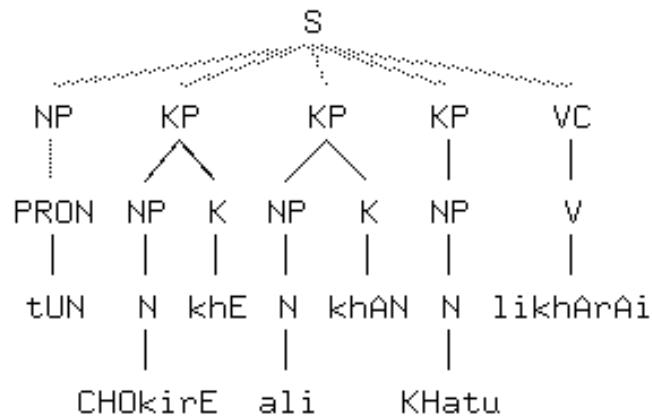
OBL: Ali

OBJ2: CHOkirO

OBJ: KHatu

Verb Subcategorization

"tUN CHOkirE khe ali khAN KHatu likhArAi"



F-structure #1

'tUN CHOkirE khe ali khAN KHatu likhArAi'

	[PRED 'likhu<[1:tUN],[35:CHOkirO],[71:Ali],[108:KHat]>'
SUBJ	[PRED 'tUN' NTYPE [NSYN pronoun] [CASE nom, GEND masc, NUM sg, PERS 2, PRON-TYPE personal]
OBJ2	[PRED 'CHOkirO' NSEM [N-CONCEPT animate] NTYPE [NSEM [COMMON count] [NSYN common] [CASE dat, GEND masc, N-FORM obl, NUM sg, PP-FORM khe]
OBL	[PRED 'Ali' NSEM [N-CONCEPT animate] NTYPE [NSEM [PROPER [PROPER-TYPE name]] [NSYN proper] [CASE agent, GEND masc, NUM sg, PERS 3, PP-FORM khan]
OBJ	[PRED 'KHat' NSEM [N-CONCEPT inanimate] NTYPE [NSEM [COMMON count] [NSYN common] [CASE nom, GEND masc, NUM sg]
GEND	masc, V-Formimperative, V-Form2 causative, VTYPE main]

Verb Subcategorization

Complement (COMP)

Ali sOchyO [ta Ahmed kela khAE thO]

ali.Nom thought [that Ahmed bananas eat be.PresAux

(↑PRED)='soch<(↑SUB) ↑COMP>'

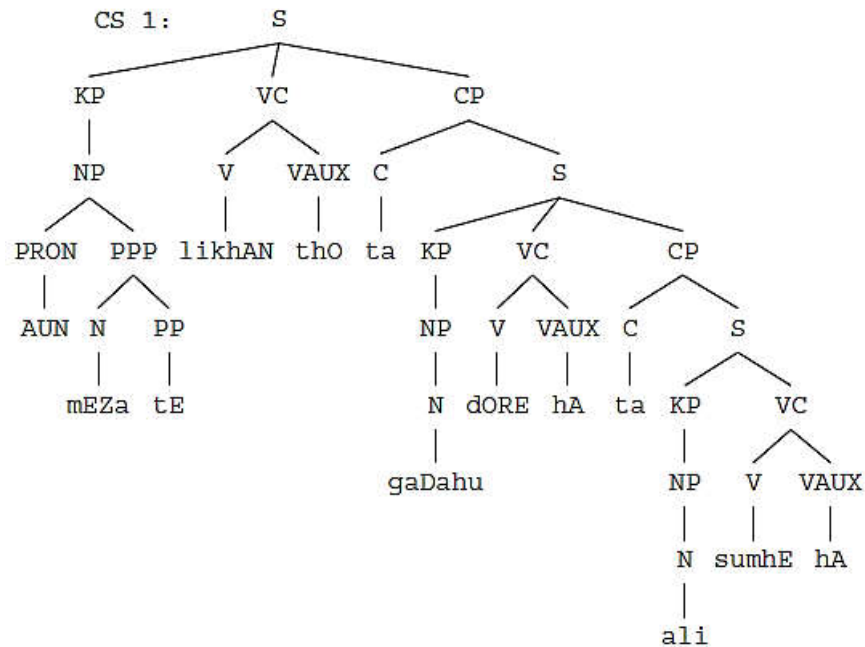
SUB: Ali

COMP: 'khau<(↑SUB) ↑OBJ>'

SUB: Ahmed

OBJ: kela

Complement (COMP)



"AUN mEzA tE likhAN thO ta gaDahu dORE hA ta ali sumhE hA"

PRED	'likhu<[1:AUN], [119:dOri]>'
SUBJ	[PRED 'AUN' ADJUNCT { [PRED 'mEzA' NSEM [N-CONCEPT inanimatē] NTYPE [NSEM [COMMON count]] [67 [CASE loc, PP-FORM on] [34 [GEND fem, N-FORM obl, NUM sg]]]]
	NTYPE [NSYN pronoun]
	1 [CASE nom, GEND fem, NUM sg, PERS 1, PRON-TYPE personal]
	[PRED 'dOri<[121:gaDahu], [187:sumhu]>' SUBJ [PRED 'gaDahu' NSEM [N-CONCEPT animatē] NTYPE [NSEM [COMMON count]] [121 [CASE acc, GEND masc, N-FORM obl, NUM sg, PERS 3]] COMP [PRED 'sumhu<[189:Ali]>' SUBJ [PRED 'Ali' NSEM [N-CONCEPT animatē] NTYPE [NSEM [PROPER [PROPER-TYPE name]]] [189 [CASE acc, GEND masc, NUM sg, PERS 3]] TNS-ASP [MOOD contrafactual, TENSE-FORM aorist] [187 [C-TYPE that, NUM sg, VTYPE main]] TNS-ASP [MOOD contrafactual, TENSE-FORM aorist] [119 [C-TYPE that, NUM sg, VTYPE main]]
TNS-ASP [MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]	
69 [AUXTYPE tho, GEND masc, NUM sg, VTYPE main]	

Verb Subcategorization

Open Complement (XCOMP)

Ali KHatu likhaNra gHure thO

Ali letter write.inf want be.AuxPres

(↑PRED)='gHuru<(↑SUBJ) (↑XCOMP)>'

SUB: Ali

XCOMP: 'kara<(↑SUBJ) ↑OBJ>'

SUB: Ali

OBJ: KHatu

"ali KHatu likhaNra ghurE thO"

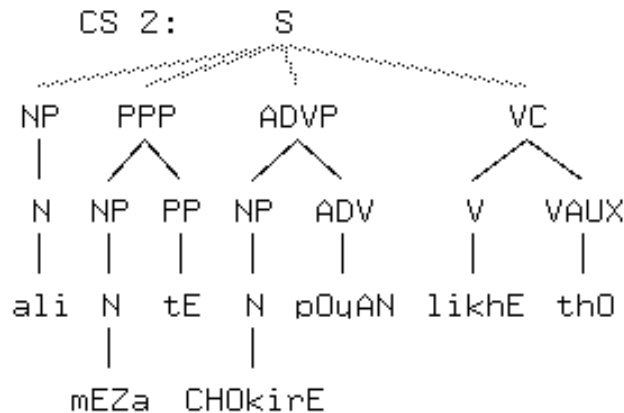
```
PRED 'ghuru<[1:Ali], [69:likhu]>'
SUBJ [
  PRED 'Ali'
  NSEM [N-CONCEPT animate]
  NTYPE [
    NSEM [PROPER [PROPER-TYPE name]]
    NSYN proper
  ]
  1[GEND masc, N-FORM obl, NUM sg, PERS 3]
]
XCOMP [
  PRED 'likhu<[1:Ali], [35:KHat]>'
  SUBJ [1:Ali]
  OBJ [
    PRED 'KHat'
    NSEM [N-CONCEPT inanimate]
    NTYPE [
      NSEM [COMMON count]
      NSYN common
    ]
    35[GEND masc, N-FORM obl, NUM sg]
  ]
  69[GEND masc, NUM pl, VFORM inf, VTYPE main]
]
TNS-ASP [MOOD indicative, PERF -, PROG -, TENSE pres.
AUXTYPE tho, GEND masc, NUM sg, VTYPE main, TENSE-FORM aorist]
```

ADJUNCT

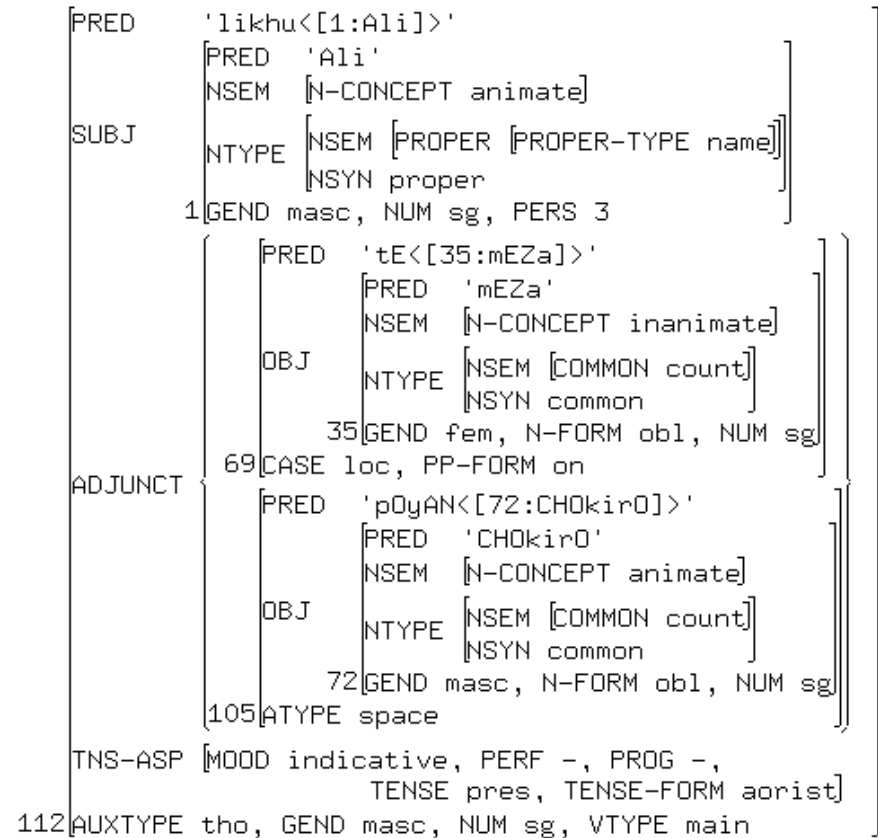
- Postpositional and adverbial phrases which do not fit in verb sub-categorization frames are called adjuncts
 - bHOlRO bAG mEN kHAE tHO
 - bHOlRO bAG mEN vaNra tE kHAE tHO
- Phrasal level Adjuncts
 - suTHO aiN suhiNrU CHOkirO

ADJUNCT

"ali mEZa tE CHOkirE pOyAN likhE thO"



"ali mEZa tE CHOkirE pOyAN likhE thO"



XAJUNCT

- XADJUNCTs are embedded sentences where SUBJ is controlled from outside
- The only pattern found is marked by conjunctive participles
 - hU dORI gHaru vayO
 - Ali kitAbu likhI mAnI kHAdHI

hU dORI gHaru vayO

Ali kitAbu likhI mAnI kHAdHI

More Research is required on XADJUNCT Patterns in Sindhi

XAJUNCT

"CHOkirO kitAbu likhI dORyO"

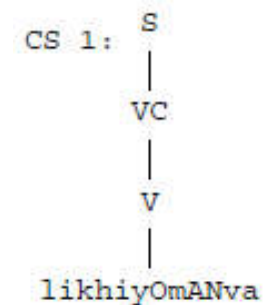
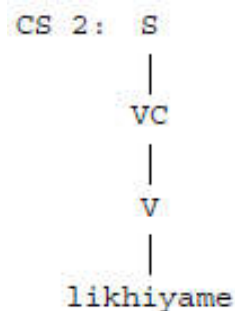
```

[PRED      'likhu<[1:CHOkirO], [34:kitAbu]>'
SUBJ      [PRED 'CHOkirO'
           NSEM [N-CONCEPT animate]
           NTYPE [NSEM [COMMON count]
                 NSYN common]
           1[CASE nom, GEND masc, NUM sg, PERS 3]
OBJ      [PRED 'kitAbu'
           NSEM [N-CONCEPT inanimate]
           NTYPE [NSEM [COMMON count]
                 NSYN common]
           34[CASE nom, GEND masc, NUM sg]
XADJUNCT [PRED 'dORi<[1:CHOkirO]>'
          SUBJ [1:CHOkirO]
          91[GEND masc, NUM sg, VTYPE main]
68[GEND masc, PTCPL-TYPE conjunctive, VTYPE main]

```

Pronominal Suffixes

Suffixes attached to verbs, construct different morphological forms, syntactically cause pro-drop



"likhiyame"

[PRED	'likhu<[1-SUBJ:pro]>']
[SUBJ	[PRED 'pro'
		[GEND fem, N-FORM obl, NUM sg, PERS 1
]]
[TNS-ASP	[PERF +
1		[NUM sg, PTCPL-TYPE past, VTYPE main
]]

"likhiyOmANva"

[PRED	'likhu<[1-SUBJ:pro], [1-OBJ:pro]>']
[SUBJ	[PRED 'pro'
		[GEND fem, N-FORM obl, NUM sg, PERS 1
]]
[OBJ	[PRED 'pro'
		[CASE acc, GEND fem, NUM pl, PERS 2
]]
[TNS-ASP	[PERF +
1		[NUM sg, PTCPL-TYPE past, VTYPE main
]]

coverage

- Morphology

- FST Models (Nouns, Pronouns, Adjectives, Verbs)
- LFG Lexicon Postpositions, Conjunctions, Adverb
- Features
 - Gender, Number, Case, Mood, Aspect, Tense

- Syntax

- Partially Free Word Order
- SUB, OBJ, OBL, OBJ2, COM, XCOMP, ADJUNCT, XADJUNCT, PREDLINK
- Coordination, Subordination, Mood, Case, Aspect, Tense, Agreement

Coverage

Word Class	Stems	Morphological Forms / Inflections	Average Inflections / Stem
Verbs	100	5013	50.13
Nouns	323	1729	5.35
Pronouns	79	283	3.58
Adjectives	71	394	5.55
Adverbs	38	38	1.00
Total	611	7457	12.20

Conclusion & Future Work

- Development in current state covers the morphological and syntactic constructions discussed in above.
- Basic morphology and syntax constructs in Sindhi are identified and modeled.
- Morphological analysis shows interesting results like adjectives have more average inflections than nouns
- Pronouns have 3.58 average inflections per word.
- Also verb can have up to 75 different morphological forms (or even more)

Conclusion & Future Work

- Though the basic constructs of Sindhi morphology and Syntax are implemented yet many complexities are subject to further research and development including:
 - pronominal suffixation with nominal elements,
 - pronominal suffixation with postpositions,
 - NP coordination model,
 - verbal complex constructions which form complex predicates,
 - Adverbial agreement
 - Prodrop phenomenon in Sindhi.

References

- [1] K. R. Beesley, and K. Lauri. "Finite-state morphology: Xerox tools and techniques." CSLI, Stanford (2003).
- [2] C. Dick, M. Dalrymple, R. Kaplan, T. H. King, John Maxwell, and Paula Newman. "XLE documentation." Palo Alto Research Center (2008).
- [3] M. U. Rahman. "Sindhi Morphology and Noun Inflections", in proc. *Conference on Language & Technology (CLT-09)*, pp. 74-81. 2009.
- [4] R. Emmanuel, and Y. Schabes. *Finite-state language processing*. MIT press, 1997.
- [5] K. R. Beesly. "Arabic Morphology Using Only Finite State Operations", in proc. *Workshop on Computational Approaches to Semetic languages*, Montreal, Quebec, pp. 50-57. (1998).
- [6] M. J. Steedman. A Generative Grammar for Jazz Chord Sequences. *Music Perception* 2 (1): 52–77. JSTOR 40285282. (1989).
- [7] M. U. Rahman., and M. I. Bhatti. "Finite State Morphology and Sindhi Noun Inflections.", in proc. *Pacific Asia Conference on Language, Information and Computation (PACLIC 24)*. Sendai, Japan. pp. 669 – 676 (2010).
- [8] M. U. Rahman, A. Shah "Grammar Checking Model for Local Languages.", in proc. *SCONEST (Student Conference on Engineering Sciences and Technology)* Karachi. (2003).
- [9] M U. Rahman, A. Shah, R.A. Memon. Partial Word Order Syntax of Urdu/Sindhi and Linear Specification Language. *Journal of Independent Studies and Research (JISR) Volume 5, Number2, July 2007.* pp. 13 – 18.
- [10] J. D. Oad. *Implementing GF Resource Grammar for Sindhi*. Unpublished Master's Thesis. Department of Applied Information Technology Chalmers University of Technology Gothenburg, Sweden. (2012).
- [11] A. Ranta. "Grammatical Framework: A Type-Theoretical Grammar Formalism", *Journal of Functional Programming* 14 (2): 145–189. (2004).
- [12] M. Butt, *The Structure of Complex Predicates in Urdu*, CSLI Publications, Stanford. (1995).
- [13] M. Butt, D. Helge, T. H. King, H. Masuichi, and C. Rohrer. "The parallel grammar project." in proc. "Workshop on Grammar engineering and evaluation" Volume 15, pp. 1-7. Association for Computational Linguistics, 2002.
- [14] T. Bögel,, M. Butt, A. Hautli, and S. Sulger. "Urdu and the modular architecture of ParGram", in proc. *Conference on Language and Technology*, vol. 70. Lahore. 2009.
- [15] S. M. J. Rizvi. *Development of Alorithms and Computational Grammar for Urdu*, PhD thesis. ept. of Computer and Information Science PIAS. Islamabad. 2007
- [16] L. Karttunen, *Finite-State Lexicon Compiler*. Technical Report, ISTL-NLTT2993-04-02, Xerox Palo Alto Research Center, Palo Alto, California (1993)
- [17] L. Karttunen, and K. R. Beesley. "Twenty-five years of finite-state morphology." *Inquiries Into Words*, a Festschrift for Kimmo Koskenniemi on his 60th Birthday (2005): 71-83.
- [18] M. Dalrymple. *Lexical-Functional Grammar*. John Wiley & Sons, Ltd, 2001.

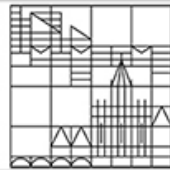
Acknowledgements

SNLP
Society for Natural Language Processing



مرکز تحقیقات لسانیات
میرزا محققا لسانیات

Universität
Konstanz



DAAD

Deutscher Akademischer Austausch Dienst
German Academic Exchange Service



CLT07, CLT09, CLT10, CLT12, CLT14, CLT16