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

<table>
<thead>
<tr>
<th>Singular</th>
<th>Intermediate</th>
<th>Plural</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRY</td>
<td>CRYS</td>
<td>CRIES</td>
<td>y→ie / ^_____s#</td>
</tr>
</tbody>
</table>

![Morphological Diagram](image)
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 (↑ PRED) = ‘mAru<(↑ SUBJ), (↑ OBJ)>’
(↑ TENNSE) = Past
(↑ SUBJ NUM) = SG
(↑ SUBJ PERS) = 3

Ali N (↑ PRED) = ‘Ali’
(↑ NUM) = SG
(↑ PERS) = 3

C-Structure Rules

1. S → NP VP
   (↑ SUBJ= ↓) ↑=↓

2. NP → N
   ↑=↓

3. VP → NP V

4. --------------
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
+Accusative +Dative +Ergative +Genitive +Instrumental + Locative +Nominative +Oblique +Vocative !Noun
Cases
+Count +Mass +Gerund +Measure +City +Country
+FirstName
+LastName
+FullName
+Name
+Fem +Masc !Gender
+Sg +Pl !Number
+1st +2nd +3rd !Person
LEXICON Root
Nouns;
LEXICON Nouns
!Boy (Animate Common Noun)
CHOkir+Noun+Common+Count+Animate:CHOkir N_Cat1;
LEXICON N_Cat1
+ChOkir+Noun+Common+Count+Animate
+Sg+Masc+Nominative

Morphological analysis of surface form “CHOkirO”

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

Mutee U Rahman, Hameedullah Kazi
Isra University, Hyderabad
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)

Tense, Aspect and Mood not yet analyzed by Sindhi Grammarians
Noun Cases

- **Case**

<table>
<thead>
<tr>
<th>Case</th>
<th>Case Marker</th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>∅</td>
<td>CHOkirO</td>
<td>CHOkirO</td>
</tr>
<tr>
<td>Accusative / Dative</td>
<td>-E</td>
<td></td>
<td>CHOkir-E</td>
</tr>
<tr>
<td>Postpositional</td>
<td>-E</td>
<td>CHOkirO</td>
<td>CHOkir-E</td>
</tr>
<tr>
<td>Locative</td>
<td>-E</td>
<td>CHOkirO</td>
<td>CHOkir-E</td>
</tr>
<tr>
<td>Instrumental</td>
<td>-E</td>
<td></td>
<td>sONT-E sAN</td>
</tr>
<tr>
<td>Possessive / Genetive</td>
<td>-E</td>
<td></td>
<td>CHOkir-E JO</td>
</tr>
<tr>
<td>Ablative</td>
<td>-AN</td>
<td>gHaru</td>
<td>gHar-AN:</td>
</tr>
<tr>
<td>Vocative</td>
<td>-A</td>
<td>CHOkirO</td>
<td>CHOkirA</td>
</tr>
</tbody>
</table>

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

Mutee U Rahman, Hameedullah Kazi
Isra University, Hyderabad
## Pronouns

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

<table>
<thead>
<tr>
<th>Case</th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom.Sg</td>
<td>kehRO: CHOKirO</td>
<td>kehRI CHOKirI</td>
</tr>
<tr>
<td>Nom.pl</td>
<td>kehRA CHOKirA</td>
<td>kehRyUN CHOKirUN</td>
</tr>
<tr>
<td>Obl.sg</td>
<td>kehRE CHOKirE</td>
<td>kehRIa CHOKirIa</td>
</tr>
<tr>
<td>Obl.pl</td>
<td>kehRani CHOKirani</td>
<td>kehRiyuni CHOKiruni</td>
</tr>
<tr>
<td>Gen.sg</td>
<td>muhiNjO CHOKirO</td>
<td>muhiNJI: CHOKirI</td>
</tr>
<tr>
<td>Gen.pl</td>
<td>muhiNjA CHOKirA</td>
<td>muhiNjUN CHOKiriUN</td>
</tr>
</tbody>
</table>
Pronominal Suffixes

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

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Pronominal Suffix Type</th>
<th>Syntactic Role</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nominal Suffix</td>
<td>Noun</td>
<td>puTa-mi</td>
</tr>
<tr>
<td>2</td>
<td>Verbal Suffix</td>
<td>Verb</td>
<td>mAri-yAN-si</td>
</tr>
<tr>
<td>3</td>
<td>Postpositional Suffix</td>
<td>Pronoun</td>
<td>kHE-na</td>
</tr>
</tbody>
</table>
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)
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
  • KPPoss is used
Verb Subcategorization

SUBJECT & OBJECT

AliWrite a book.

(↑ PRED)=‘LIKHU<(↑ SUB) (↑ OBJ)>’

"AUN kitAbu likhan thO"

Mutee U Rahman, Hameedullah Kazi

Isra University, Hyderabad
Verb Subcategorization

SUBJECT Only

\[ \text{Ali} \; \text{dORE} \; \text{tHO} \]

Ali runs.
(↑PRED)=’dORi<(↑SUB)’
Verb Subcategorization

Passives: SUBJ $\rightarrow$ NULL, OBJ $\rightarrow$ SUBJ

kitAbu likHijE thO
Book is being written/Book writing takes place.

($\uparrow$PRED)='LIKHU<(NULL $\uparrow$SUB)>'

"kitAbu likHijE"

kitAbu likHibO AhE
Book writing takes place.

($\uparrow$PRED)='LIKHU<(NULL $\uparrow$SUB)>'

Mutee U Rahman, Hameedullah Kazi

Isra University, Hyderabad
Passives: NULL Arguments

likHibO AhE
Writing takes place.
(↑PRED)=’LIKHU<(NULL)>’

likHijE tHO
(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

(↑PRED)=‘likhu<(↑SUB) (↑OBJ2) (↑OBJ)>’
SUB: ali
OBJ2: CHOkirO
OBJ: KHatu
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

\[
\begin{array}{c}
\text{PRED 'likhu([1:tUN],[35:CHOKir0],[71:Ali],[108:KHat])'} \\
\text{PRED 'tUN'} \\
\text{SUBJ NTYPE [NSYN pronoun]} \\
\text{CASE nom, GEND masc, NUM sg, PERS 2, PRON-TYPE personal} \\
\text{PRED 'CHOKir0'} \\
\text{NSEM [N-CONCEPT animate]} \\
\text{OBJ2 NTYPE [NSYN [COMMON count]]} \\
\text{NSYN common} \\
\text{CASE dat, GEND masc, N-FORM obl, NUM sg, PP-FORM khe} \\
\text{PRED 'Ali'} \\
\text{NSEM [N-CONCEPT animate]} \\
\text{OBL NTYPE [NSYN [PROPER [PROPER-TYPE name]]]} \\
\text{NSYN proper} \\
\text{CASE agent, GEND masc, NUM sg, PERS 3, PP-FORM khan} \\
\text{PRED 'KHat'} \\
\text{NSEM [N-CONCEPT inanimate]} \\
\text{OBJ NTYPE [NSYN [COMMON count]]} \\
\text{NSYN common} \\
\text{CASE nom, GEND masc, NUM sg} \\
\text{GEND masc, V-Form imperative, V-Form2 causative, VTYPE main}
\end{array}
\]
Verb Subcategorization

Complement (COMP)


(↑PRED)='soch<(↑SUB)↑COMP>'

SUB: Ali

COMP: ‘khau<(↑SUB)↑OBJ>’

SUB: Ahmed

OBJ: kelA
Complement (COMP)
Verb Subcategorization

Open Complement (XCOMP)

Ali KHatu likhaNra gHuruE thO

Ali letter write.inf want be.AuxPres

(↑PRED)=’gHuru<(↑SUBJ) (↑XCOMP)>’

SUB: Ali

XCOMP: ‘kara<(↑SUBJ) ↑OBJ>’

SUB: Ali

OBJ: KHatu
• 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
"ali mEZa tE CHOkirE pOyAN likhE thO"

CS 2: S
NP PPP ADVP VC
| | | |
N NP PP NP ADV V VAUX
| | | |
ali N tE N pOuAN likhE thO
| | |
mEZa CHOkirE

TNS-ASP [MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]
112 [AUXTYPE tho, GEND masc, NUM sg, VTYPE main]
• 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

More Research is required on XADJUNCT Patterns in Sindhi
"CHOkir0 kitAbu likhI dORy0"

[ PRED 'likhu<[1:CHOkir0], [34:kitAbu]>'
  [ PRED 'CHOkir0'
    [ NSEM [N-CONCEPT animate]
      SUBJ [NTYPE [NSEM [COMMON count]
        NSYN common
        1CASE nom, GEND masc, NUM sg, PERS 3
      ]
    ]
    OBJ [NSEM [N-CONCEPT inanimate]
      NTYPE [NSEM [COMMON count]
        NSYN common
        34CASE nom, GEND masc, NUM sg
      ]
    ]
  ]
  XADJUNCT SUBJ [1:CHOkir0]
  91GEND masc, NUM sg, VTYPE main
  68GEND masc, PTCPL-TYPE conjunctive, VTYPE main ]
Pronominal Suffixes

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

<table>
<thead>
<tr>
<th>Word Class</th>
<th>Stems</th>
<th>Morphological Forms / Inflections</th>
<th>Average Inflections / Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs</td>
<td>100</td>
<td>5013</td>
<td>50.13</td>
</tr>
<tr>
<td>Nouns</td>
<td>323</td>
<td>1729</td>
<td>5.35</td>
</tr>
<tr>
<td>Pronouns</td>
<td>79</td>
<td>283</td>
<td>3.58</td>
</tr>
<tr>
<td>Adjectives</td>
<td>71</td>
<td>394</td>
<td>5.55</td>
</tr>
<tr>
<td>Adverbs</td>
<td>38</td>
<td>38</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>611</td>
<td>7457</td>
<td><strong>12.20</strong></td>
</tr>
</tbody>
</table>
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

Acknowledgements

CLT07, CLT09, CLT10, CLT12, CLT14, CLT16