

**DEVELOPING A SINDHI COMPUTATIONAL
RESOURCE GRAMMAR IN LEXICAL
FUNCTIONAL GRAMMAR FRAMEWORK**

**A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy**

by

Mutee U Rahman

**Department of Computer Science
Faculty of Engineering Science and Technology
Isra University, Hyderabad**

January 2017

DEVELOPING A SINDHI COMPUTATIONAL RESOURCE GRAMMAR IN LEXICAL FUNCTIONAL GRAMMAR FRAMEWORK

by

Mutee U Rahman

Names of Supervisor and Co- Supervisors

**Dr. Hameedullah Kazi (Supervisor)
Professor (Computer Science)**

**Dr. Ahsanullah Baloch (Co-Supervisor)
Professor (Computer Science)**

**Dr. Muhammad Iqbal Bhatti
Professor (Computer Science)**

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. Hameedullah Kazi, whose encouraging guidance, constructive criticism, and advice always motivated me during this research study.

I am also thankful to Tafseer Ahmed Khan who guided me at every step of my research from topic selection to thesis writing. His kind support, encouragement and company always invigorated me throughout this research work.

I also owe sincere gratitude to Miriam Butt who was always there whenever I needed any help in learning LFG, XLE and linguistic concepts. I am thankful for her detailed review with valuable comments. Specially her support for attending LFG conferences is unforgettable for me as it was an amazing learning experience.

I would also like to express my sincere gratitude to Rajesh Bhatt for his encouraging criticism, advice and comments.

I am thankful to Martin Frost who taught me LFG and XLE along-with Miriam Butt during ISSALE 2014 at University of Colombo. Discussions with him helped me a lot to understand LFG basics.

I am also thankful to Sebastian Lorry for his valuable time to help in debugging while LFG grammar development during my visits at University of Konstanz Germany.

I have great regard for Dr. Sarmad Hussain who always discussed with me about my research and kept encouraging me about my research work. I also have great regard for Dr. Ghulam Ali Allana who helped me to understand

some basic concepts of Sindhi language.

My sincere thanks also go to my co-supervisors Dr. Ahsanullah Baloch and Dr. Muhammad Iqbal Bhatti for guiding me throughout this research work.

I am also grateful to Hafiz Tariq Ali Shaikh for valueable discussion on different linguistic topics which provided basis for understanding different Sindhi language constructs.

I also wish to express my sincere gratitude to Society for Natural Language Processing, Pakistan (SNLP) and German Academic Exchange Service (DAAD) who provided me oppurtunities to attend International Summer School in Advanced Language Engineering (ISSALE 2014) and to visit University of Konstanz, Germany during my research work.

Last but not least loving gratitude goes to my wife and children who always supported me beyond their capacity.

ABSTRACT

Computational grammar development and deep linguistic analysis provides structural details for natural language understanding by machines. Modern multilingual information processing systems use these details for understanding and processing of information represented in different languages. While work in Sindhi language is focused in the areas like part of speech tagging and machine learning. Sindhi lacks resources like computational grammars and deep linguistic analysis systems. Development of such resources is open research area in computational linguistic and natural language processing domains.

This work presents the development of Sindhi language morphology and grammar in Finite State Technology and Lexical Functional Grammar (LFG) frameworks. The work includes the investigation and identification of morphology and syntax patterns in Sindhi language, development of Sindhi finite state lexicon by modeling of identified morphological patters in LEXC, development of Sindhi LFG by incorporating the finite state lexicon in XLE, and evaluation of developed morphological lexicon and LFG grammar.

Various parts of speech of Sindhi language are investigated and their morphological patterns are identified. Nouns are marked by number, gender and case. Ten different cases of nouns are identified namely nominative, accusative, dative, participant, instrumental, locative, ablative, agentive, genitive and vocative. Adjectives are also declined like nouns. Pronouns are declined for number and gender and are marked by nominative, oblique and genitive cases. Generally, adverbs are not inflected but when adjectives used as adverbs they hold the inflectional properties of adjectives. Genitive

postpositions are inflected and marked by number and gender. Conjunctions and interjections do not inflect. Verbs are most complex part of speech and classified into main, auxiliary, copula and modal verbs. Verbs are conjugated by number and gender and are marked by tense, aspect and mood. Morphological analysis of developed model shows that a verb can have up to 75 different morphological forms in Sindhi. Present, past and future tense patterns along with aspect and mood are analyzed. Aspect in Sindhi can either be perfective or imperfective (continuous and habitual) and can be marked morphologically or syntactically. Many alternative patterns of different aspects exist. Nine different mood patterns are identified including subjunctive, presumptive, imperative, declarative, permissive, prohibitive, capacitive, compulsive and suggestive. Pronominal suffixes in Sindhi may appear on nouns, postpositions and verbs. Pronominal suffixation can possibly cause subject and object pro-drop.

Sindhi syntax is analyzed with LFG perspective. Different noun phrase constructions are implemented with coordination patterns including adjective phrases, postpositional phrases, participle phrases, and relative clauses. Genitive case marking patterns along with syntactic agreement are identified and modeled in LFG. Verbal subcategorization frames are defined for different grammatical functions including SUBJ (Subject), OBJ (Object), OBJ2 (Secondary Object), OBL (Oblique), COMP (Complement), XCOMP (Open Complement), and PREDLINK (Predicate link). Phrase and sentence level adjuncts (ADJUNCT) and open adjunct (XADJUNCT) patterns are also identified and implemented in LFG.

The developed grammar is tested against two different test suites. First

test suite contains 617 handcrafted sentences in 10 different test files containing sentences with different syntactic features. Second test suite contains real time corpus of two text books of Sindhi class one with 258 sentences. Results show 98.05% and 96.5% parsing percentage of test suite 1 and test suite 2 respectively.

Morphology coverage includes 862 stems of different POS classes with total of 10327 inflectional forms. The developed finite state morphology is tested and evaluated against the corpus of 9050 words in terms of coverage, ambiguity, precision, recall and f-measure (F_1). The results show 97.8% precision, 96.08% recall and average ambiguity of 1.65 solutions per word with 91.1% coverage. Coverage of different morphological features include number, gender, case, tense, aspect and mood. Syntactic coverage includes nominal elements, coordination, subordination, agreement, verbal subcategorization, tense, aspect and mood.

Research and development results include Sindhi part of speech tagset, roman script for Sindhi language, morphological lexicon and LFG grammar of Sindhi. As a side development, a corpus of about 4 million words is also developed. In absence of linguistic resources for Sindhi language, these developments will have signification impact on Sindhi language processing and further research in computational linguistics and related domains.

ABBREVIATIONS & SYMBOLS

CFG	Context Free Grammar
CLE	Center for Language Engineering
CRBLP	Center for Research on Bangla Language Processing
CRULP	Center for Research on Urdu Language Processing
EAGLES	Expert Advisory Group on Language Engineering Standards
FOPC	First Order Predicate Calculus
FSA	Finite State Automata
FSM	Finite State Machine
FST	Finite State Transducer
GWB	Grammar Writer's Workbench
HPSG	Head Driven Phrase Structure Grammar
IE	Information Extraction
IR	Information Retrieval
LFG	Lexical Functional Grammar
NLP	Natural Language Processing
POS	Part of Speech
PSG	Phrase Structure Grammar
TAG	Tree Adjoining Grammar
XLE	Xerox Linguistic Environment
XTAG	eXtended Tree Adjoining Grammar

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	i
ABSTRACT	iii
ABBREVIATIONS & SYMBOLS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xvii
CHAPTER I	1
INTRODUCTION	1
1. BACKGROUND	2
2. PROBLEM STATEMENT	5
3. OBJECTIVES AND INTENDED OUTCOMES	6
4. WHY FINITE STATE MORPHOLOGY AND LEXICAL FUNCTIONAL GRAMMAR?	6
5. GRAMMAR ENGINEERING MODEL.....	8
6. TRANSLITERATION SCHEME.....	8
7. ORGANIZATION OF THE THESIS.....	9
CHAPTER II	11
LITERATURE REVIEW	11
1. FINITE STATE MORPHOLOGY	12
2. LEXICAL FUNCTIONAL GRAMMAR	14
2.1 Formal Definition of Lexical Functional Grammar	15
2.1.1 Lexicon and Argument Structure.....	15
2.1.2 Constituent structure (c-structure).....	17
2.1.3 Functional Structure (f-structure)	18
3. RELATED WORK	20
3.1 Related Work in Finite State Morphology.....	20
3.2 Related Work in Syntax Modeling within LFG Framework	23
4. RELATED WORK FOR SINDHI.....	25
CHAPTER III.....	28

METHODS AND TOOLS	28
1. SINDHI WORD CLASSES	28
1.1 Closed Word Classes in Sindhi.....	29
1.1.1 Pronouns	30
1.1.2 Postpositions.....	31
1.1.3 Conjunction.....	32
1.1.4 Auxiliary, Copula and Modal Verbs.....	32
1.1.5 Uncategorized function words.....	32
1.1.6 Interjection	33
1.2 Open Word Classes in Sindhi	33
1.2.1 Nouns	33
1.2.2 Adjectives	34
1.2.3 Adverbs.....	34
1.2.4 Verbs and Compound Verbs.....	35
1.2.5 Participles	36
2. MORPHOLOGY	37
2.1 Sindhi Words.....	37
2.2 Morphological Constructions in Sindhi	38
2.3 Noun Morphology.....	39
2.3.1 Gender.....	39
2.3.2 Number.....	40
2.3.3 Case	41
2.4 Adjective Morphology.....	45
2.5 Pronoun Morphology	48
2.6 Postposition Morphology.....	53
2.7 Adverb Morphology	53
2.8 Conjunction and Interjection Morphology	55
2.9 Sindhi Verb Morphology	55
2.9.1 Imperatives	55
2.9.2 Second Person Imperative (The Root).....	57

2.9.3 Polite Imperative	57
2.9.4 Desiderative	58
2.10 Formation of Infinitive	59
2.11 Formation of Participles	59
2.11.1 Present participle	59
2.11.2 Past participle	60
2.11.3 Future participle	61
2.11.4 Verbal Noun	62
2.11.5 Conjunctive Participle	63
2.12 Transitive and Causative Formations	64
2.13 Tense and Aspect	65
2.13.1 Aorist (Old Present) formation with Unspecified Aspect.....	66
2.13.2 Present Tense.....	67
2.13.3 Past Tense.....	68
2.13.4 Future Tense.....	70
2.13.5 Perfective Aspect.....	73
2.13.6 Imperfective-Continuous Aspect	76
2.13.7 Imperfective-Habitual Aspect	79
2.13.8 Combining Aspects	80
2.14 Mood	83
2.14.1 Subjunctive Mood	83
2.14.2 Presumptive Mood.....	85
2.14.3 Imperative Mood	86
2.14.4 Declarative or Indicative Mood.....	86
2.14.5 Permissive, Prohibitive, Capacitive, Suggestive and Compulsive Moods	87
2.15 Pronominal Suffixes	89
2.15.1 Double Pronominal Suffixation.....	93
2.16 Postpositional Suffixes	95
3. SYNTAX.....	96

3.1 Noun Phrase in Sindhi	96
3.1.1 Complex NP Constructions	98
3.1.2 Postpositional Phrases and Relative Clauses	101
3.2 Case Marking	103
3.2.1 Nominative Case	103
3.2.2 Obliqueness and Ergativity	104
3.2.3 Accusative Case	105
3.2.4 Differential Object Marking	106
3.2.5 Dative Case	107
3.2.6 Participant Case	108
3.2.7 Instrumental Case	109
3.2.8 Locative Case	109
3.2.9 Ablative Case	110
3.2.10 Agentive Case	110
3.2.11 Possessive / Genitive Case	111
3.2.12 Case Marking with Case Phrase	113
3.3 Verb Sub Categorization	117
3.3.1 Subject	117
3.3.2 Object	120
3.3.3 OBJ2 (Secondary Objects)	121
3.3.4 Oblique	121
3.3.5 Complement and Open Complement	122
3.3.6 Adjunct	123
3.3.7 Open Adjunct (XAdjunct)	124
4. IMPLEMENTATION	124
4.1 Implementing Sindhi Morphology (Simple Word Classes)	125
4.1.1 Sindhi Noun Morphology	125
4.1.2 Sindhi Adjective Morphology	130
4.1.3 Sindhi Pronoun Morphology	133

4.1.4 Sindhi Postposition Morphology	139
4.1.5 Sindhi Adverb Morphology	141
4.1.6 Sindhi Conjunction and Interjection Morphology	143
4.2 Implementing Sindhi Morphology (Verbs and Pronominal Suffixes)..	143
.....	
4.2.1 Auxiliary and Copula Verbs	145
4.2.2 Verbs.....	151
4.2.3 Pronominal Suffixes	155
4.3 Implementing Syntax in XLE	162
4.4 XLE Grammar File Structure	162
4.5 Implementing Nominal Syntax.....	164
4.5.1 Nouns in Noun Phrase.....	164
4.5.2 Pronouns in Noun Phrase.....	166
4.5.3 Adjectives in Noun Phrase	170
4.5.4 Noun Phrases with Postpositional Phrases	172
4.5.5 Case Marking of Nouns / Noun Phrases.....	174
4.5.6 Case Phrase (KP)	176
4.5.7 Possessive / Genitive Case	181
4.6 Implementing Verbal Syntax	187
4.6.1 Subject.....	187
4.6.2 Object	190
4.6.3 Secondary Object (Obj2)	192
4.6.4 Oblique	195
4.6.5 Complement (COMP)	196
4.6.6 Open Complement (XCOMP)	199
4.7 Adjunct	201
4.8 Open Adjunct (XAdjunct).....	203
4.9 Relative Clause and NP with Relative Clause.....	205
4.10 Pronominal Suffixes	208
CHAPTER IV	210

RESULTS	210
1. EVALUATION OF FINITE STATE MORPHOLOGY.....	210
2. TEST SUITES AND LFG EVALUATION	214
2.1 Description of Test Suite-1	214
2.1.1 Test File 1	215
2.1.2 Test File 2	216
2.1.3 Test File 3	217
2.1.4 Test File 4	218
2.1.5 Test File 5	219
2.1.6 Test File 6	221
2.1.7 Test File 7	222
2.1.8 Test File 8	224
2.1.9 Test File 9	226
2.1.10 Test File 10	227
2.2 Test Suite 2	228
3. RESULTS.....	231
3.1 Unparsed Sentences.....	233
4. GRAMMAR COVERAGE	234
5. OTHER OUTCOMES OF RESEARCH AND DEVELOPMENT..	237
CHAPTER V.....	239
DISCUSSION	239
CHAPTER VI	249
CONCLUSIONS AND FUTURE WORK.....	249
REFERENCES.....	257
APPENDIX A	267
SINDHI PART OF SPEECH TAGSET	267
APPENDIX B	268
ROMAN SCRIPT FOR SINDHI.....	268

LIST OF TABLES

Chapter	Description	Page
II-1: Sample orthographic rule.....		13
III-1: Closed word classes in Sindhi.		29
III-2: Sindhi Pronouns and their subclasses.....		30
III - 3: Types of Sindhi Postpositions.....		31
III - 4: Conjunction types in Sindhi.....		32
III - 5: Types of Sindhi Nouns.....		33
III - 6: Examples of adjectives in Sindhi.		34
III - 7: Adverb types in Sindhi.		35
III - 8: Sindhi Compound Verbs.		36
III - 9: Participles in Sindhi.		36
III - 10: An example of FST rule.		39
III - 11: Sindhi Masculine and Feminine Nouns.....		40
III - 12: Number inflections in feminine and masculine nouns		41
III - 13: Noun Case Marking Inflections of Masculine Nouns.....		46
III - 14: Noun Case Marking Inflections of Feminine Noun CHOKirla “girl” ..		46
III - 15: Sindhi vocative case examples.....		47
III - 16: Declension of Native Adjective “suTHO”.....		47
III - 17: Adopted Adjective without Declension.....		47
III - 18: Personal Pronoun Declensions.....		49
III - 19: Demonstrative Pronoun Declension Patterns		50
III - 20: Wh-Pronoun Declension Patterns.....		51
III - 21: Relative and Co-Relative Pronoun Declension Patterns.....		52
III - 22: Declension of Indefinite Pronouns.		52
III - 23: Postposition Morphology.....		53
III - 24: Adverb Inflections		54

III - 25: Verb Roots.....	57
III - 26: Polite and more-polite imperative formation.....	58
III - 27: Desiderative formation	58
III - 28: Number marking on imperatives.....	58
III - 29: Formation of infinitives	59
III - 30: Formation of Present Participle.....	60
III - 31: Formation of Past Participle.....	61
III - 32: Formation of Future Participle.....	62
III - 33: Formation of Verbal Noun.....	63
III - 34: Conjunctive Participle Formation	64
III - 35: Intransitive to transitive formation	65
III - 36: Formation of Aorist	67
III - 37: Present Tense Auxiliary Conjugation.....	68
III - 38: Present Tense Formation Examples.....	68
III - 39: Transitive verb “likHu” with past indefinite tense.....	69
III - 40: Number and Gender Suffixes for Past Tense Verbs.....	69
III - 41: Past Tense Suffixes for Intransitive Verbs	70
III - 42: Past Tense Conjugation Examples of Intransitive Verb dOri	71
III - 43: Future Tense Suffixes.....	71
III - 44: Future Tense Morphology of Transitive Verb “likHu”	72
III - 45: Future Tense Morphology of Intransitive Verb “dORi”	72
III - 46: Future Tense Morphology of Auxiliary “AhE”	73
III - 47: Perfective Aspect Patterns in Different Tenses by using Perfective Participle	74
III - 48: Perfective Aspect Patterns with Conjunctive Participle and Completion Marker Auxiliary	76
III - 49: Imperfective Continuous Constructions with “payO” Marker	77

III - 50: Imperfective Continuous Constructions with Conjunctive Participle and “rah-iyO” marker.....	78
III - 51: Imperfective Habitual Construction	79
III - 52:Continuous-Habitual Construction	81
III - 53: Alternate Continuous-Habitual Construction	82
III - 54: Subjunctive Mood Constructions	84
III - 55: Presumptive Mood Formations	85
III - 56: Imperative Mood Forms	86
III - 57: Permissive, Prohibitive, Capacitive, Suggestive, and Compulsive Mood Constructions	88
III- 58: Pronominal Suffixes Used with Nouns	89
III - 59: Pronominal Suffixes used with Present Tense Auxiliary “AhE”	91
III- 60: Pronominal Suffixes of Past Tense Auxiliary “hu-O”	91
III- 61: Pronominal Suffixes in Old-Present/Aorist Verb Forms	92
III - 62: Pronominal Suffixation on Postpositions.....	93
III - 63: Double Pronominal Suffixation Examples of Verb “liKhu” 1 st Person – 2 nd Person and 1 st Person – 3 rd Person	94
III - 64: Double Pronominal Suffixation Examples of Verb “liKhu” 3 rd Person – 1 st Person, 3 rd Person – 2 nd Person and 3 rd Person – 3 rd Person.....	94
III - 65: Postpositional Suffixes	95
III - 66: Oblique forms of common noun ‘CHOkirO’	104
III - 67: Different Morphological Forms of Common Noun Stem “CHOkir” .	128
III - 68: Morphology of Native Adjective “suhiNrO” with Intermediate and Surface Forms	132
III - 69: Morphology of Native Adjective “thOrO” and Borrowed Adjective “KHUbsUrat” with Intermediate and Surface Forms.....	132
III - 70: Morphology of Adjective “hika” with Intermediate and Surface Forms	133
III - 71: Intermediate and Surface Forms of First Person Pronoun Generated by Figure III-7 Entries.....	136

III - 72: Intermediate and Surface Forms of Second Person Pronoun Generated by Figure III-7 Entries.....	137
III - 73: Different Forms of Auxiliary “AhE” Generated by Figure VI-10 Transducer.....	148
III - 74: Intermediate and Surface Forms of “tHO” and “payO”	150
III - 75: Different Morphological Forms of Transitive Verb “likHu”.....	154
III - 76: Pronominal Suffix Surface Forms Generated by Pronominal Suffix Morphology	157
VI - 77: Different Morphological Forms of Intransitive Verb “dORi”.....	160
VI - 78: Different Morphological Form of Intransitive Verb “dORi”	161
IV - 1: Stems and Surface Forms Coverage.	211
IV - 2: Inflectional Categories and Average Inflections.....	211
IV - 3: Results of Morphological Analysis.....	214
IV - 4: Comparison with Apertium Implementation	214
IV - 5: Unparsed Sentences with Explanation.....	235
IV - 6: Grammar Coverage.....	235
IV - 7: Completely Used LFG Rules.	236
IV - 8: Partially Used LFG Rules.	237
B - 1: Partially Used LFG Rules.	269

LIST OF FIGURES

Chapter	Description	Page
I - 1:	Grammar Development and Testing Model.....	8
II - 1:	Two level morphology process using FSTs.....	13
II - 2:	Sample lexical entry in LFG.....	16
II - 3:	LFG Lexicon.....	17
II - 4:	C-structure rules for a simple Sindhi sentence.....	18
II - 5:	C-structure tree of simple Sindhi sentence.....	18
II - 6:	General form of f-structure in LFG.....	19
II - 7:	Correspondence between c-structure and f-structure.....	20
III - 1:	Possessive Noun Phrase Examples.....	116
III - 2:	Incorrect Possessive Noun Phrase Parse.....	116
III - 3:	Thematic Hierarchy to Identify Grammatical Roles.....	119
III - 4:	Sindhi Noun Morphology LEXC Snapshot.....	127
III - 5:	Proper Noun Entries in LEXC	129
III - 6:	Adjective Morphology Entries	131
III - 7:	Pronoun Morphology Entries in LEXC	134
III - 8:	Different Pronoun Types in LEXC Morphology	138
III - 9:	Full Form Lexicon Entries for Postpositons.....	140
III - 10:	Adverbs Lexicon Entries in LEXC	142
III - 11:	Full Form Lexicon Entries for Conjunctions	143
III - 12:	Multi_Char Symbols Used in Verb Morphology	145
III - 13:	Auxiliary AhE Morphology in LEXC.....	147
III - 14:	Morphology of Auxiliaries “tHO” and payO in LEXC	149
III - 15:	Lexicon Entries of Transitive Verb “likHu” and Intransitive Verb “dOri” in LEXC	151
III - 16:	Verbstem1 Sublexicion Definition	152

III - 17: Infinitive and Passive Ending Sublexicon Definitions	153
III - 18: Psuffix (Pronominal Suffix) Sublexicon Definitions	156
III - 19: Sublexicon VerbStem2 (Intransitive Verbs) Definition	159
III - 20: Sindhi LFG Grammar File Structure	163
III - 21: Parse Tree and F-structure of Noun “CHOkirO”	165
III - 22: F-structure Analyses of Different Forms of Common Noun “CHOkirO”	165
III - 23: Noun Phrase Definition with Pronominal Elements in LFG	166
III - 24: LFG Analysis of Different Personal Pronouns.....	167
III - 25: NP construction with demonstrative and Wh-pronouns	168
III - 26: NP construction with Reflexive and Indefinite Pronouns.....	169
III - 27: Simple ADJP (Adjective Phrase) Analysis	171
III - 28: Coordinated ADJP Analysis.....	171
III - 29: NP Analysis with Three Coordinated Adjectives.....	172
III - 30: An NP with Postpositional Phrase	173
III - 31: Parse Tree of Complex NP Construction Using Postposional having Embedded NP.	174
III - 32: F-structure Analysis of Complex NP Construction Using Postposional Phrase having Embedded NP.....	175
III - 33: Case Analysis of Common Noun Form “CHOkirA”	176
III - 34: Accusative and Dative Case Marking of Common Noun “CHOkirO”	177
III - 35: Participant and Instrument Case Marking with Case Marker “sAN”.	178
III - 36: Locative Case Marking with “tE” and “mEN” Markers.....	179
III - 37: Ablative and Agentive Case Marking.....	180
III - 38: Morphological Genitive Case Marking of First and Second Person Pronouns.....	182
III - 39: Plural Genitive Case Marking with Marker “jA”.	183

III - 40: NP Example with Possessive Specifier.....	184
III - 41: Inconsistant F-structure Due to Different Genders of Maker and Possessed Noun.....	185
III - 42: Nested NP Example.	185
III - 43: F-structure Analysis of Nested NP.....	186
III - 44: LFG Analysis of Sample Sentence “CHOkirO dORE”.....	189
III - 45: F-structure Analysis of Passive Verb “dORijE”.	189
III - 46: Transitive Verb likhu with “SUBJ” and “OBJ” Subcategorization....	191
III - 47: Verb “likhu” with SUBJ Subcategorization Only.	192
III - 48: Passivization of of “likhu” with NULL Subject formation.....	192
III - 49: Parse Tree for a Sentence with SUBJ, OBJ2, and OBJ.....	194
III - 50: F-structure Analysis of a Sentence with SUBJ, OBJ2, and OBJ....	194
III - 51: Passive Formation with SUBJ, OBJ2, and OBJ.....	195
III - 52: Parse Tree of a Sentence with SUBJ, OBJ2, OBL, and OBJ Arguments.....	196
III-53: F-structure Analysis of a Sentence with SUBJ, OBJ2, Agentive OBL and OBJ Subcategorization.....	197
III - 54: Sentence with Embedded Complement Phrase (CP).	198
III - 55: F-structure Analysis of a Sentece with COMP Argument.....	198
III - 56: Parse Tree of a Sentence with Embedded XCOMP Phrase.....	200
III - 57: F-structure Analysis a Sentence with SUBJ and XCOMP Subcategorization.....	200
III - 58: F-structure Analysis a Sentence with SUBJ, XCOMP and OBJ2 Subcategorization	201
III - 59: Sentence with Adjuncts.....	202
III - 60: F-structure Analysis a Sentence with XCOMP Subcategorization.	203
III - 61: Sentence with XADJUNCT.	204
III - 62: F-structure Analysis of a Sentence with XADJUNCT.....	204
III - 63: Parse Tree of a Relative Clause	206

III - 64: F-structure Analysis of a Relative Clause.....	206
III - 65: F-structure Analysis of an NP with Relative Clause.....	207
III - 66: F-structure Analysis of verb “likhu” with Pronominal Suffix.....	209
IV - 1: C-structure of a sentence with PREDLINK and Copula Verb.....	215
IV - 2: F-structure Analysis of a sentence with PREDLINK and Copula Verb.	215
IV - 3: C-structure of a sentence with Coordinated Adjectives PREDLINK.	216
IV - 4: F-structure of a sentence with Coordinated Adjectives PREDLINK.	216
IV - 5: C-structure of a sentence with Compound NP Construction.....	217
IV - 6: F-structure Analysis of a Sentence with Compound NP Construction.	218
IV - 7: C-structure of a sentence with Passive Verb.....	219
IV - 8: F-structure Analysis of a Sentence with Passive Verb.	219
IV - 9: C-structure of a sentence with Phrase level and Sentence Level Adjuncts.....	220
IV - 10: F-structure Analysis of a Sentence with Phrase and Sentence Level Adjuncts.....	220
IV - 11: C-structure of a sentence with SUBJ and XCOMP.....	221
IV - 12: F-structure Analysis of a Sentence with SUBJ and XCOMP Subcategorization.....	222
IV - 13: C-structure of a Sentence with Syntactically Formed Genitive Case.	223
IV - 14: F-structure Analysis of a Sentence with Syntactically Formed Genitive Case.....	223
IV - 15: C-structure of a Sentence with Nested Complement Phrase.	224
IV - 16: F-structure Analysis of a Sentence with Nested Complement Phrase.	225
IV - 17: C-structure of a Sentence with Pro-dropped Subject.	226
IV - 18: F-structure Analysis of a Sentence with Pro-dropped Subject.....	226
IV - 19: C-structure of a Sentence with Object Pro-drop Caused by Pronominal Suffixation.	227

IV - 20: F-structure Analysis of a Sentence with Object Pro-drop Caused by Pronominal Suffixation.....	227
IV - 21: Snapshot of a Page from Sindhi Nursery Book and Transliterated Text Used in Test File 1 of Test Suite 2.....	228
IV - 22: Snapshot of a Page from Sindhi Book One and Transliterated Text Used in Test File 2 of Test Suite 2	228
IV - 23: C-structure of a Sentence from Test File 1 of Test Suit 2 with Pro- dropped Subject, and XADJUNCT.....	229
IV - 24: F-structure Analysis of a Sentence from Test File 1 of Test Suit 2 with Pro-dropped Subject, and XADJUNCT.....	229
IV - 25: C-structure of a Sentence from Test File 2 of Test Suit 2 with Sentence Level Predicative Adverb ADJUNCT.....	230
IV - 26: F-structure Analysis of a Sentence from Test File 2 of Test Suit 2 with Sentence Level Predicative Adverb ADJUNCT.....	231
IV - 27: Parsing Results of Individual Files of Test Suite 1.....	232
IV - 28: Parsing Results of Individual Files of Test Suite 2.....	232
IV - 29: Parsing Results of Individual Test Suites.	233
IV - 30: Consolidated Parsing Results of Test Suite 1 and 2.	233

CHAPTER I

INTRODUCTION

Computational linguistics (CL) synonymously called Natural Language Processing (NLP) is a discipline between linguistics and computer science which focuses the on the processing of human languages on computers. NLP is an active research area not only in academia but in commercial application development as well (Agerri, et al., 2015) (Kenneth & Lisa, 1995). The ultimate goal of NLP research is to process, understand and reproduce written or spoken human languages. Lexical analysis is the first step involved in written human language processing and deals with the structure of words in a language. In other words, lexical analysis deals with the morphology of a language. Words are combined to form phrases and sentences in a language. Syntax analysis (also called parsing) deals with sentence structures and how words are combined to make phrases and sentences. Other tasks involved in NLP include semantic analysis which deals with meaning, pragmatics which deals with relationship between meaning, goals and intensions, and discourse which deals with linguistic units larger than single utterance or sentence. NLP components are essential part of modern multilingual information processing systems and require linguistic-knowledge of languages being used or processed. This knowledge is made available through linguistic research and development of the formal models of various linguistic constructs spanning from morphology to discourse. Most popular formal models used for this purpose include Finite State Automata (Roche & Shabes, 1997), Context Free Grammars (Chomsky, 1956), Lexical Functional Grammar (LFG) (Bersnan,

1982, 2001), Head Driven Phrase Structure Grammars (Levine & Detmar, 2006), Grammatical Framework (Ranta, 2004), and First Order Predicate Calculus (Barwise, 1977).

This research study investigates and implements the morphology and syntax patterns of Sindhi language (the word ‘Sindhi’ refers to ‘Sindhi language’ henceforth) in Finite State Morphology and LFG frameworks by using XFST LEXC and Xerox Linguistic Environment (XLE). The study also explores the less studied topics of Sindhi morphology and syntax including pronominal suffixes, tense, aspect, mood, and verbal subcategorization.

Subsequent sections discuss background, problem statement, objectives and intended outcomes, reason of selecting LFG and finite state models, grammar engineering model and outline of the dissertation.

1. BACKGROUND

Modern information systems incorporate different NLP components to process language specific information more effectively. These components use linguistic knowledge at different levels discussed above. This study focuses on the development and evaluation of morphology and syntax of Sindhi language in XFST LEXC and LFG frameworks respectively.

Morphology studies various patterns of word formation within a language. Finite state networks are well known computational devices to capture and represent morphological knowledge of a language. The idea of modeling morphology in finite state transducers has its roots in finite state properties of phonological rules (Johnson, 1972). Later, Koskenniemi

(Koskenniemi, 1983) presented finite state rules for Finnish and Antworth (Antworth, 1990) worked on two level morphology of English. Kaplan and Kay (Kaplan & Kay, 1994) presented regular models for computational phonology. These models became foundations for phonological and morphological processing of natural languages. Additional algorithms and mathematical foundations were given by Mohri (Mohri, Pereira & Riley, 2002), and Roche and Schabes (Roche & Schabes, 1997). These finite state transducers first model the morphological information at different levels and are combined in such a way that the output of one layer becomes the input of next layer and thus either morphological analysis takes place or surface form word is generated. Various NLP components like morphological lexicon/dictionary, phonological and orthographical rules can be implemented in finite state networks (Beesley & Karttunen, 2003).

Syntax deals with sentence level constructs in a language. Syntax analysis or parsing is a key research area in NLP studies. Context free grammar (CFG) also known as phrase structure grammar (Bakus, 1960) is one of the most popular syntax representation formalisms. However, CFG in its original form has problems of over-generation while dealing with natural language syntax constructs like agreement, sub categorization and movement (Wintner, 2002). Other formal grammars based on CFG solved most of the problems with CFGs. Tree adjoining grammars (TAG) (Vijay & Joshi, 1991) and lexicalized TAG (XTAG, 1995) are such extensions in which trees are used instead of rules. Head driven phrase structure grammar (HPSG) (Levine & Detmar, 2006) and LFG (Dalrymple, M., 2001) are most prominent extended formalisms based on CFGs which are now considered complete grammar

frameworks in formal linguistics. HPSG and LFG both are based on lexicalist approach which means that lexicon is not just a list of words but a very well structured rich lexicon. Also, both formalisms are unification and constraint based grammars (Shieber, 1986). LFG deals with syntax at different levels of representation and successfully extracts and represents language dependent and language independent information of syntax which plays important role in machine translation, syntax analysis and other applications. Linguistic depth and efficient parsing of LFG grammar makes it an attractive choice for grammar modeling and development. LFG uses finite state morphology discussed above with richly formatted lexicon and represents the syntactic relations at various levels. LFG research group is organizing regular LFG conferences since last 19 years and online proceedings of conferences are also made available (CLSI, 2016). Computer implementations of LFG grammar are available for more than 30 different languages (Sulger, et. al, 2013) including Arabic (Attia, 2008), English (Isa, 2003), Turkish (Zelal & Kemal, 1994), Bangla (Haque & Khan, 2005) and Urdu/Hindi (Hussain, S., 2004b). Any language specific syntax implementation in LFG makes that language part of LFG family. Moreover, computational modeling of any language syntax makes available the knowledge of that language to be used by researchers working in NLP, IR, linguistics and other related domains.

Sindhi syntax representation efforts in Context Free Grammars and Linear Specification Language can be found in (Rahman & Shah, 2003) and (Rahman, Shah and Memon, 2007). First study tries to represent selected Sindhi sentence structures in CFG rules and has over generation problems. Second study tries to cope with over generation by using LSL (Linear

Specification Language) but again lacks the agreement problem solution and feature representations. The only comparatively comprehensive research study available is “Implementing GF Resource Grammar for Sindhi” (Oad, 2012). This study tries to investigate the Sindhi morphology and syntax from computational perspective in grammatical framework (Ranta, 2004). Basic syntactic and morphological constructions are implemented without going into the details of complex linguistic constructions; a Sindhi resource grammar library is made available as a shared resource. Sindhi however still lacks linguistic resources to be used by language processing systems and is subject of research within computational linguistics domain.

2. PROBLEM STATEMENT

Sindhi is a resource poor language in Computational Linguistics and Natural Language Processing domains. Neither Sindhi morphology nor the syntax is studied by researchers with computational linguistics perspective. Rich inflectional constructions of Sindhi morphology with number, gender, case, tense, aspect, mood and pronominal suffixation need to be analyzed and developed in modern linguistic frameworks. Sindhi syntax features including free constituent ordering, agreement, complex noun phrase constructions, coordination, subordination, syntactic case formations, pro-drop and verbal subcategorization are open research areas. Sindhi morphological and syntactic constructions therefore need to be investigated and modeled in suitable linguistic frameworks to create usable linguistic resources for modern information processing systems and related domains.

3. OBJECTIVES AND INTENDED OUTCOMES

Objective of this research study are to:

- 1) investigate and classify Sindhi Morphology patterns
- 2) investigate and classify Sindhi Syntax patterns
- 3) develop Sindhi Finite State Morphology
- 4) develop Morphology tags and POS (part of speech) tagset for Sindhi
- 5) develop Sindhi LFG Grammar
- 6) evaluate Sindhi morphology and LFG Grammar

The intended outcomes of the research include:

- 1) Theoretical foundations for Sindhi NLP and Computational Linguistics
- 2) Morphological lexicon for Sindhi
- 3) Sindhi syntax analyzer in LFG framework
- 4) Sindhi POS tagset

4. WHY FINITE STATE MORPHOLOGY AND LEXICAL FUNCTIONAL GRAMMAR?

The descriptive power of finite state networks along-with their mathematical properties and operations made possible the availability of tools for working on morphological modeling of languages. These tools include Xerox XFST, LEXC, and TWOC (Kenneth & Lauri, 2002). By using these tools morphological analyzers for various languages are designed and evaluated

which include English (Cohen, et al., 2003), French (Chanod, 1994), German (Gunther & Dietrich, 1988), Spanish (Manuel et al., 1996), Dutch (Antal & Walter, 1999), Urdu (Hussain, S., 2004a) (Bogel, et. al, 2007) and others.

XFST can compile huge FST networks with millions of paths in seconds and thus FST lexicon network for morphological analysis and generation is made available. Morphological modeling and lexicon development in finite state networks using Xerox tools makes it easy to analyze and evaluate the morphological modeling of Sindhi language. Cross lingual morphological analysis of Sindhi and other languages for which finite state network models are already designed is also possible. Moreover, XFST developed morphology can be integrated easily in LFG via morphology syntax interface (Kaplan & Butt, 2002) within XLE.

LFG framework is chosen for proposed research due to its popularity, maturity, robustness, handling of syntactic issues like free word order and complex predicates which occur frequently in south Asian languages, and language independent nature. LFG research groups regularly hold LFG conferences (LFG, 2016) in which researchers participate from different countries working on different languages.

Research on Sindhi morphology, lexicon and syntax within the frameworks of FSM and LFG will produce valuable literature and resources for researchers working on NLP in these frameworks in particular and computer science and linguistics in general.

5. GRAMMAR ENGINEERING MODEL

The overall research and development model is shown in Figure I-1. Based on linguistic survey and research Sindhi morphology is investigated and developed within finite state models using XFST, Lexical Functional Syntax and LFG based lexicon are developed by using XLE. XLE also interfaces between morphology and syntax components. As a result, Sindhi LFG grammar is developed. Sindhi sentences are given as input to LFG grammar and resulting parse trees and functional structures are generated.

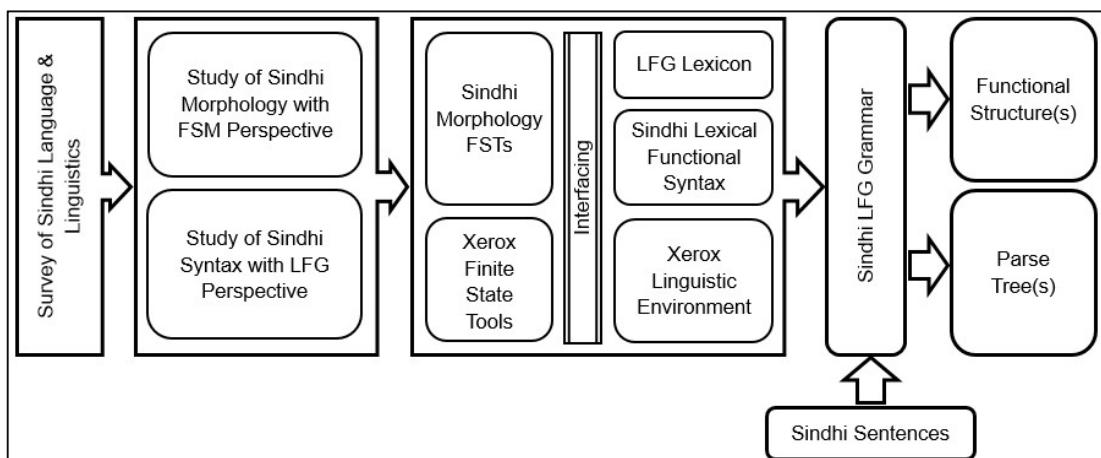


Figure I - 1: Grammar Development and Testing Model.

6. TRANSLITERATION SCHEME

A Romanized transliteration scheme is used in this research study instead of original Persio Arabic script of Sindhi language. The objective behind this decision was to make available this research to international research community in a more understandable manner as Romanized script will be in line with English translations in large number of examples discussed in this study. The roman script used throughout the research study is designed

on almost same guidelines on which Urdu roman script is used in ParGram project (Kamran, et al., 2010). Complete description of the roman transliteration scheme used is given in Appendix B.

7. ORGANIZATION OF THE THESIS

Rest of the thesis proceeds as follows. Chapter II is about literature review which discusses the classical work in the fields of finite state morphology and LFG with their formal definitions. Existing research work in closely related languages and Sindhi is also discussed. Chapter III discusses methods and tools. First section discusses Sindhi word classes in detail along-with their closed and open class grouping. Second section discusses morphology of simple and complex word classes, declension patterns of simple word classes including nouns, pronouns, adjectives, postpositions, adverbs, conjunctions and interjections are discussed with number, gender and case marking. Then, comparatively complex word class verb and its morphology is discussed, various morphological forms of different types of verb are discussed with tense, aspect and mood formations. Discussion in Chapter III continues with pronominal suffixes. After that Chapter III discusses Sindhi syntax in detail. Different nominal elements and constructions of noun phrase are explored with its constituent parts. Syntactic case marking of noun phrase with different case markers, verbal syntax and sub-categorization is also discussed in this chapter. Various arguments of clause including SUBJ, OBJ, OBJ2, OBL, COMP, XCOMP, and PREDLINK are explored and analyzed. Sentence and phrase level adjuncts are also discussed in Chapter III. Last section of Chapter III contains the implementation details of Sindhi

morphology and syntax in LEXC and XLE. LFG implementation in XLE is also given in this section. This section further explores the implementation details of LFG with verbal subcategorization. Results are presented in Chapter IV. Discussion is given in Chapter V and finally conclusion and future work is given in Chapter VI followed by references and appendices.

CHAPTER II

LITERATURE REVIEW

Natural Language Processing (NLP) deals with computational techniques and models that represent and process spoken and written human languages. Modeling and processing of a language can be at various levels discussed in chapter I; however, scope of this study is limited to morphological and syntactic analysis of Sindhi in Finite State Morphology and LFG frameworks by using XFST and Xerox Linguistics Environment respectively.

Models of morphological analysis always remained challenge for computational linguists until early 1980's when 4Ks¹ discovered the two-level morphology (Beesley & Lauri, 2003) the first general model for morphologically complex languages. This two-level morphology represents a word at lexical and surface levels. Morphotactics or morpheme ordering models are placed in between to handle morphological variations between lexical and surface levels. These morphotactics are implemented as Finite State Transducers (FSTs) which consume one level as input and produce other level as output.

Syntactic analysis of a language requires grammar construction of the language in a suitable syntax representation framework. Popular syntax representation models include Context Free Grammars (Hopcroft & Ullman, 1979), Tree Adjoining Grammars (Vijay & Weir, 1994), Head Driven Phrase Structure Grammars (Carl & Ivan, 1994), Grammatical Framework (Ranta, 2004) and LFG (Falk, 2001). Among all these formal models, LFG is extensively studied and used for South Asian languages particularly for

¹ Ronald M. Kaplan, Martin Kay, Lauri Karttunen and Kimmo Koskenniemi

Urdu/Hindi (Butt & King, 2007) (Sulger & Vaidya, 2014).

Related work in NLP domain mostly include the research studies for content dominant languages of the world including English (Stanford, 2016a), Chinese (Stanford 2016b), Russian (NLP Center, 2012), Hindi (IIT, 2016), Urdu (Anwar & Wang, 2006) and many more. Few preliminary studies are also available in the related areas of Sindhi NLP discussed in subsequent sections.

Following sections discuss finite state morphology, LFG, research work in closely related languages and existing work for Sindhi.

1. FINITE STATE MORPHOLOGY

Finite state transducers play an important role in language processing applications (Roche & Shabes, 1997) and computational studies of morphologically complex languages. As discussed earlier morphotactics (morpheme ordering rules) are represented by finite state transducers; as a result, efficient morphological parsers are implemented.

Finite state transducers convert/translate lexical level constructs to surface level words by applying morphotactics. Their reversible nature makes reverse conversion/translation possible. These two levels (lexical and surface) morphology plays essential role in implementation of morphological analyzers for natural languages. Figure II-1 shows the process of two level (lexicon and surface) morphology modeling using FSTs. Figure II-1 (a) shows the finite state transducer where either upper or lower layer is used as input and the other one as output. A sample orthography FST rule is given in Table II-1. The substitution rule “y→ie / ^____s#” rule says that “y” will be replaced with “ie” whenever it is between morpheme boundary “^” and ending “s”, it may be noted

that “^” and “#” are used as a morpheme boundary and word boundary respectively. This rule simply converts intermediate plural forms with “-ys” ending into “-ies” as shown in Table II-1. Orthography FSTs use this rule to convert intermediate level words into surface words. Overall conversion process can be seen in Figure II-1 (b). Figure II-1 (c) shows the block diagram of this process.

Table II-1: Sample orthographic rule.

Singular	Intermediate	Plural	Rule
CRY	CRY S	CRIES	y → ie / ^ ____ s#

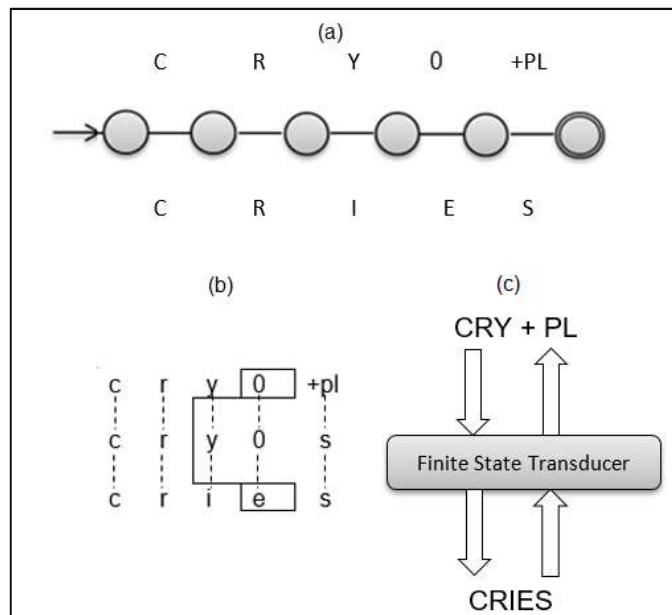


Figure II - 1: Two level morphology process using FSTs.

Finite state models based on these FSTs are well known and successfully been used for morphological modeling of many languages. They handle concatenative and non-concatenative morphology very well (Beesely, 1998b). As discussed above FSTs are reversible i.e. if they run in reverse order they generate the surface level words this can be seen in Figure II-1 (c). FSTs therefore not only work as morphological analyzers but surface word

generators as well.

2. LEXICAL FUNCTIONAL GRAMMAR

LFG is a natural language syntax representation formalism based on generative grammars (Steedman, 1989). LFG defines the linguistic structure and relationship among them. Various relations are defined at lexicon level as LFG has a rich lexical structure. As the name explains LFG is lexical (i.e. with rich lexicon) and it is functional in a sense that basic grammatical functions like subject, object, indirect object, and oblique are at the core of LFG theory.

LFG represents linguistic structures at different levels which include lexicon, constituency structure (c-structure) and functional structure (f-structure) levels. The lexicon contains list of words or parts of words (morphemes) along-with information about these words including their syntactic and morphological features. Thus, a lexical entry in LFG may include part of speech, number, gender, case, and argument structure (in case of verbs, some adjectives and postpositions). Syntactic structure information in LFG is represented at two different levels. C-structure representation handles word or phrase grouping and their precedence in a phrase structure tree along-with some grouping and order constraints. F-structure represents more abstract relations between different functional constructs like SUBJ (subject), OBJ (object), and OBL (oblique) etc. Section 2.1 discusses formal definition of LFG and explores different components of LFG framework.

2.1 Formal Definition of Lexical Functional Grammar

LFG defines structural representation of a language at three different levels (Sells, 1985). In LFG framework, these multiple levels are defined independently with their respective syntax and architecture; they represent different linguistic aspects of a language with specific formal properties (Nordlinger, 1998). These levels are linked with each other through the principles of correspondence (Nordlinger & Bersnan, 2011) which ensures the compatibility of different linguistic constructs.

Multiple levels of linguistic representation in LFG include Lexicon along-with argument structure (a-structure), constituent structure (c-structure), functional structure (f-structure), semantic structure (σ -structure), phonological structure and thematic structure. Last three levels deal with logical and semantic interpretation, sound system, pragmatics and discourse (Austin, 2001). First three are relevant to syntactic analysis and are discussed in following sections.

2.1.1 Lexicon and Argument Structure

In LFG, lexicon is not only a list of words but it encodes many syntactic properties as well (Pienemann, 2005). Lexicon entries contain information about c-structure category of each word as well as its functional structure. By doing this LFG shunts most of the explanatory burden on the lexicon and away from grammar rules (Carroll, 2007). Therefore, lexicon requires special attention in LFG. Formally LFG lexicon description contains at least three entries listed below (Dalrymple, 1999):

- Representation of the form of the item
- Syntactic category of the item
- List of functional schemata

Sample lexical item entry is shown in Figure II-2.

Here lexical item is surface of the word i.e. how it appears in the sentence where syntactic category is part of speech class of the word and list of functional schemata is list of features i.e. information about the lexical entry. The metavariable “ \uparrow ” denotes the parent category (item in this case) followed by feature name and value; “(\uparrow NUM) = SG” therefore specifies that item “Ali” has singular number.

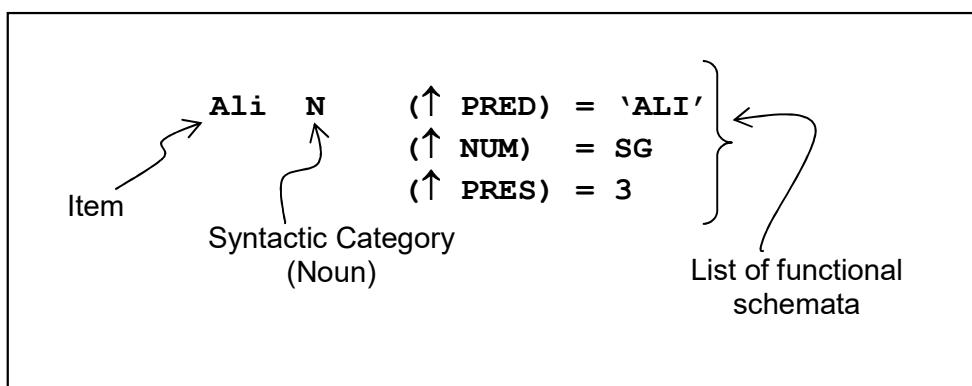


Figure II - 2: Sample lexical entry in LFG.

This information may also contain argument structure if there are arguments in the lexical entry like in case of verbs. Figure II-3 shows lexicon entries in which the entry for verb “mAryO” (killed) says that “mAryO” is a verb with root form “mAru” and has two obligatory arguments SUBJ and OBJ. It also says that verb “mAryO” is a past tense form. Next it says that subject must be third person singular form. These lexical entries bring information to the sentence when items are inserted into c-structure level (discussed below).

mAryO	V	(↑ PRED)	=	'mArū<(↑ SUBJ), (↑ OBJ)>'
		(↑ TENNSE)	=	Past
		(↑ SUBJ NUM)	=	SG
		(↑ SUBJ PERS)	=	3
Ali	N	(↑ PRED)	=	'Ali'
		(↑ NUM)	=	SG
		(↑ PERS)	=	3

Figure II - 3: LFG Lexicon.

2.1.2 Constituent structure (c-structure)

Constituent structure in LFG represents word order and phrasal grouping rules. These rules define that how words in sentences are grouped in constituents. Formally constituent structure in LFG is defined as extended context free grammar rules with functional schemata on the right-hand side of the rule. These functional schemata (also called functional equations) are in the form of special annotations which define the relationship between c-structure and f-structure (discussed in next section). A sample c-structure rule is shown in Figure II-4.

Rule 1 in Figure II-4 says that a sentence (S) is defined as sequence of noun phrase (NP) followed by a verb phrase (VP). Additional functional equations are given below syntactic categories NP and VP. In these equations, the symbol (\uparrow) known as meta variable and represents the image or f-structure of parent node and (\downarrow) represents f-structure of current node (node itself). The equation (\uparrow SUBJ) = \downarrow below NP in rule 1 states that subject in the f-structure of parent node S is f-structure of current node NP. In other words, NP contains

the subject of the sentence. The equation $\uparrow=\downarrow$ states that f-structure of child node is same like the f-structure of parent node. C-structure of sample Sindhi sentence is given in Figure II-5.

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

Figure II - 4: C-structure rules for a simple Sindhi sentence.

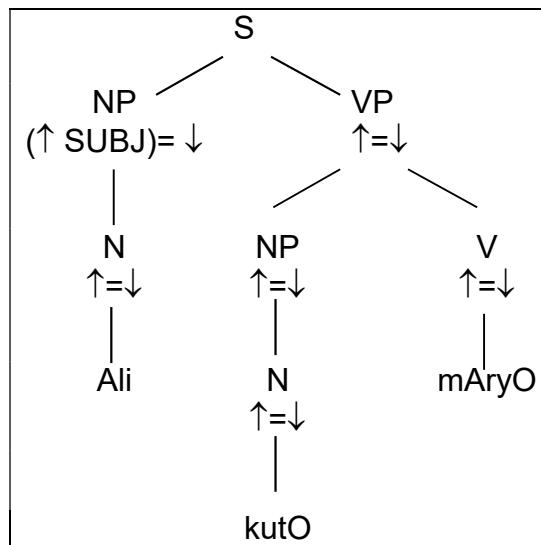


Figure II - 5: C-structure tree of simple Sindhi sentence.

2.1.3 Functional Structure (f-structure)

Functional Structure (also called feature structure) represents grammatical relations by using attribute-value structure known as AVM (Attribute Value Matrix). Each f-structure in LFG contains two column entries enclosed in large square brackets. First column entry is known as attribute and second column entry is called value (together AVM). Attribute values can be

simple attribute value pairs or nested f-structures of set of values. Figure II-6 shows general form of f-structure in LFG.

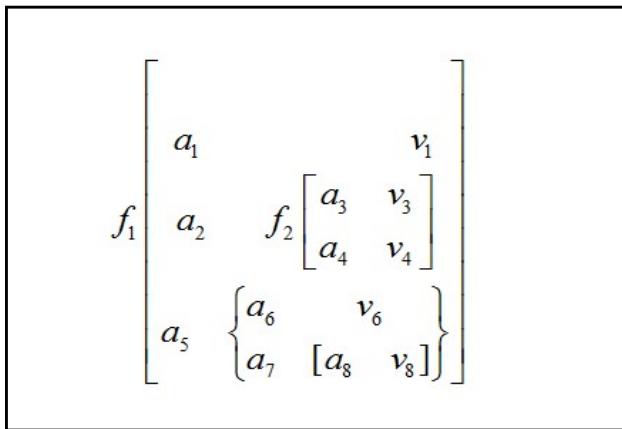


Figure II - 6: General form of f-structure in LFG.

The attributes a_1 , a_3 , a_4 and a_6 in figure II-6 have simple values (it may be noted that f_n , a_n , and v_n represent feature n, attribute n and value n respectively). The attribute a_2 has another f-structure as a value. The attribute a_5 has a set of values. F-structures are generally considered language independent representation of syntactic relations (Rosen, 1996). F-structures are constructed by using lexical entries, their attributes along-with c-structure rules and functional equations. Each f-structure must satisfy following three conditions (Nordlinger, 1998):

- i) *Uniqueness*: Attribute of an f-structure can have at most one value. However different attributes may have same values (Bersnan, 2001).
- ii) *Completeness*: F-structure must define all attributes referred in the sentence.
- iii) *Coherence*: Every defined grammatical function must be referred.

The correspondence between c-structure and f-structure of a sample Sindhi sentence “ali kutO mAryO” (Ali killed the dog) is shown in Figure II-7.

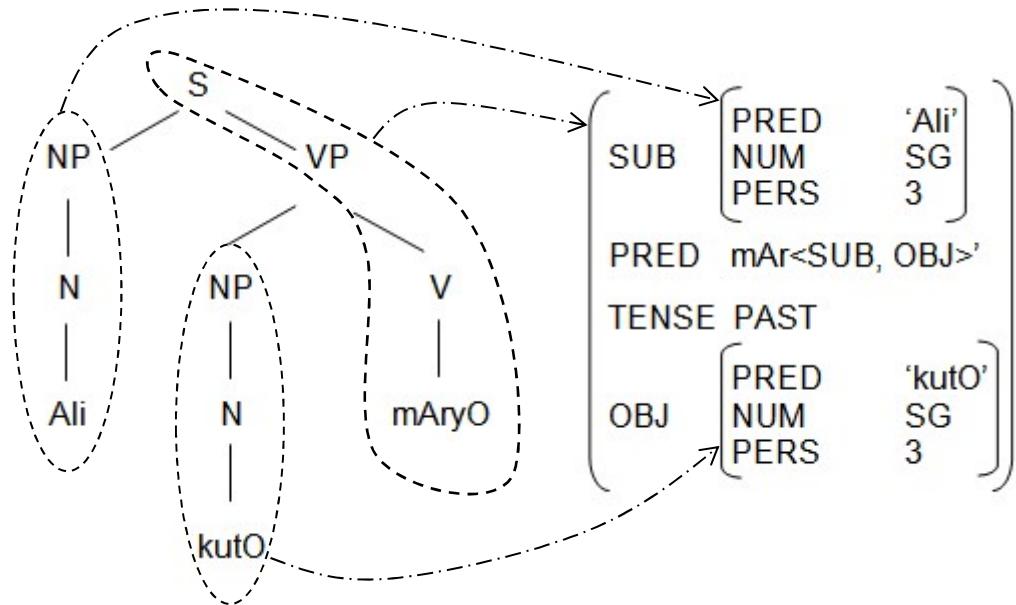


Figure II - 7: Correspondence between c-structure and f-structure.

3. RELATED WORK

Research studies within finite state morphology using Xerox FST and LFG framework using XLE exist for different languages. Various language specific and general issues are addressed by different researchers with NLP perspective. Related work in Arabic, Persian Urdu/Hindi and Punjabi is discussed in subsequent sections.

3.1 Related Work in Finite State Morphology

Research in Arabic finite state morphology started in early 1990s. studies well known in literature include (Beesely, 1996), (Beesely 1998a), (Bukwater, 2002) and (Attia et, al. 2011). First study is about Arabic morphological analysis and generation. This study investigates different challenges in Arabic morphological constructions and proposes a model based

on XFST. The language specific challenges discussed for Arabic include identification of word like prefix morphemes such as conjunctions “wa” and “fa”, prefixed prepositions such as “bi” and “li”, definite article, verbal prefixes and suffixes, enclitic direct object, and passive pronoun suffixes. Also, the problem of identification of roots is discussed. The study concludes that despite of difficulties Arabic morphology is fully susceptible to finite state morphology models. Buckwater morphological analyzer is also well known in literature. Different word forms are represented in a database along-with separate concatenation lists. It is less ambiguous than Xerox Arabic morphological analyzer; however, it does not capture the generalities as word forms are entered manually which also increases the cost of maintenance. Also, it is not suitable for generation. Muhammad Attia in his PhD thesis (Attia, 2008) addresses the morphological and syntactic ambiguities in Arabic in LFG framework using XLE. Morphological ambiguities in Arabic due to complex morphological constructions are resolved by using stem as the base form instead of root. A rule based system based on finite state morphology is proposed. Muhammad Attia further worked on Arabic morphology with special forms in lexicon development and acquisition for morphological analysis (Attia et, al., 2011). Multitier structure of Arabic morphology along-with morphotactics is discussed and lemma based approach is proposed and developed.

Persian morphological studies in finite state morphology framework are also in place. Finite state morphological analysis of Persian (Megerdoomian, 2004) presents two level morphological analyzer using XFST. Morphological modeling issues like long distance morphological dependencies and word-morpheme boundary detection problems are discussed. Problems of detached

inflections, complex tokens are solved by using finite state calculus. PerLex (Sagot & Walther, 2010) a morphological lexicon for Persian is another morphological resource available. Persian morphology is discussed and analyses show that comparatively simple morphological construction are there in Persian than Arabic.

Finite state morphology modeling of Urdu in Xerox tools was firstly developed as a part of Pargram project (Butt, et. al, 1999). In their research paper “Urdu and Parallel Grammar Project” (Butt & King, 2002) they discussed the finite state morphology development process for Urdu as a two step process. In first step, morphological classes, their sub-types and tags are identified. Secondly morphological patterns are identified and modeled. Later Sara Hussain in her research study (Hussain, S., 2004a) discussed various topics on computational morphology of Urdu. Verbal and nominal morphology is discussed and implemented; however, implemented parser efficiency is questionable. (Bogel, et. al, 2007) investigates and discusses morphological analyzer for Urdu. Different untouched issues of reduplication and echo duplication are also discussed and modeled by using two level morphology in Xerox tools. Few morpho-syntactic issues are also highlighted. Jafar Rizvi in his PhD thesis (Rizvi, 2007) also discussed various morphological classes comprehensively. The study investigates Urdu morphology in finite state morphology framework. Sixty (60) basic forms of Verbs along-with infinitive, perfective, repetitive, subjunctive and imperative forms are discussed. Classification of Urdu case markers is also given.

3.2 Related Work in Syntax Modeling within LFG Framework

A comprehensive research study in known literature is “Handling Arabic Morphological and Syntactic Ambiguities with the LFG framework with a View to Machine Translation” (Attia, 2008). Different syntactic ambiguities in Arabic are addressed and resolved in LFG framework. Other issues addressed include: free word order, pro-drops, and complex sentence structures. The study proposes an Arabic LFG parser for machine translation system. Agreement, functional control and long distance dependencies are discussed and modeled in LFG. The study also discusses the grammar development, testing and evaluation processes. Earlier Attia in his research work on multiword expression in Arabic (Attia, 2006) worked on handling multiword expressions at various levels of LFG. There is no direct way to deal with multiword expression due to possible inflections and passivization problems. It is therefore shown that these multiword expressions of Arabic can be modeled at different levels in LFG including lexicon, syntax and semantics.

Though many studies are in progress in LFG framework for Persian but one can find only few studies in existing literature. The research study “Application of LFG theory in Designing an English Persian Machine Translation” (Farough & Jahangeri, 2008) discusses the ambiguities in translation by using semantic information through lexical mapping. Later (Ghochani, 2010) discusses syntactic constructs of Persian with an LFG approach. The study investigates light verbs and complex predicates in Persian syntax with LFG implementation. Another research study (Asuder & Mortaza vinia, 2011) discusses obligatory control in Persian along-with its

implication on syntax semantics interface. A formal constraint based analysis with LFG perspective is provided. Mismatch is identified at syntax semantics interface and resolved through flexible typological semantic composition in glue semantics. A lexical functional model for machine translation of English zero place predicates into Persian is presented by (Ahanger et. al., 2012). It is shown that zero place predicates of English are translatable in Persian (however they don't exist in Persian). The study shows that zero place predicates translation by using LFG based machine translation is more natural and is based on Persian word order.

Among South Asian languages Urdu is extensively studied with LFG perspective. Miriam Butt (Butt, M., 1995) in her PhD thesis on Urdu complex predicates identified and analyzed complex predicates in Urdu. Complex predicates previously were considered as compound or composite verbs. She gave her own definition of complex predicates and examined permissive and aspectual complex predicates in Urdu with LFG perspective. A unifying theory of complex predicate formation was formulated and different issues related to argument structure linking as case marking were addressed. It was shown that presented complex predicates theory is also applicable to Romance and Japanese. Later, Urdu became part of parallel grammar project (Butt, et., al, 1999), (Butt & King, 2002) and analyzed with large scale grammar development perspective. It was found that basic analysis decisions made for European languages are applicable to typologically different language Urdu. In ParGram project parallel computational grammar of different languages is being developed within LFG framework. Various research articles discussing different syntactic issues in Urdu LFG including complex predicates, clitics,

argument structure, argument scrambling in noun phrases and verb phrases can be found on official website of Urdu pargram (Urdu Pargram, 2016). Jafar Rizvi in his PhD thesis (Rizvi, 2007) presented Urdu syntax analysis in LFG. The study investigates Urdu nominal and verbal elements in detail. Modeling of Urdu nominal syntax with case markers and postpositions is presented. It is shown that how various cases are represented at different levels of LFG. Verb argument structures and their modeling in LFG is also discussed. LFG lexicon and c-structure rules for Urdu are defined and presented.

Apart from studies discussed above research within finite state morphology and LFG is being conducted for various languages (Sulger, et. al, 2013) including French (Clement & Kinyon, 2001), German (Stefanie, 2003) (Lionel, et al., 2002), Pashto (Bogel T., 2010) and Bangla (Haque & Khan, 2005). Other research studies using number of different platforms are also in place which include development of Urdu resource grammar (Humayoun, 2006) and development of Punjabi resource grammar (Humayoun & Ranta, 2010).

4. RELATED WORK FOR SINDHI

Existing work on Sindhi morphology include two studies, a GF resource grammar for Sindhi (Oad, 2012) with morphological analyzer and a finite state morphological analyzer for Sindhi using Apertium ' Ittoolbox (Motlani, et al., 2016). The GF analyzer includes around 360 lexical entries of different part of speech classes. Evaluation results of morphological analysis are not reported; however, 97% accuracy is reported in terms of translation of sentences. Apertium finite state morphological analyzer is first ever openly available

morphological analyzer for Sindhi with nominal and verbal morphology coverage. Though the coverage in terms of number of stems is good as compared to GF analyzer (3454 versus 361) with 72 paradigms but some uncommon figures are presented; for example, 66 stems for postpositions are reported; postpositions are closed word class and in Sindhi only few postposition stems are there. The results are evaluated in terms of precision, recall, coverage and mean ambiguity. 97.68% precision and 97.52% recall against gold standard of 384 forms is achieved with known tokens and 97.68% precision and 72.61% recall is achieved with all tokens. Average mean ambiguity of 3.3% with 78% coverage is reported.

To the best of our knowledge literature about Sindhi syntax analysis in modern linguistic frameworks like LFG is not available; however, Sindhi syntax representation efforts in Context Free Grammars and Linear Specification Language can be found in (Rahman & Shah, 2003) and (Rahman, M. et. al, 2007). As discussed earlier in Chapter I first study has over generation problems and second study lacks the agreement problem solution and feature representations. GF Resource Grammar for Sindhi (Oad, 2012) also includes syntax coverage along-with morphology where a preliminary framework for morphology and syntax of Sindhi is presented; however complex morpho-syntactic features of Sindhi are still subject to research in GF.

Few computational linguistics resources are also available which include an online dictionary (CLE, 2016), and a POS tagset (Mahar & Memon, 2010a). Some preliminary NLP research studies for Sindhi are also in place which include part of speech tagging (Mahar & Memon, 2011), (Mahar & Memon, 2010b), (Mahar et al., 2011), and text to speech modeling (Mahar et

al., 2010). Recently various online dictionaries are made available by Sindhi Language Authority (SLA, 2016).

CHAPTER III

METHODS AND TOOLS

This chapter discusses the methods, tools and techniques used in linguistic research while developing finite state morphology and LFG grammar. First, Sindhi word classes and their characteristics along-with subtypes are discussed in detail. Next, comprehensive discussion on morphological patterns of different Sindhi word classes is presented. Different morpho-syntactic patterns identified during research are discussed in detail. Then, syntax of Sindhi is presented and discussed with nominal and verbal constructions with LFG perspective. Finally, implementation details of morphology and LFG syntax in LEXC and XLE are given.

1. SINDHI WORD CLASSES

Sentences and phrases are composed of words. Words belong to different categories called part-of-speech classes and every class has its own characteristics, morphological pattern and syntactic functions. Arrangement of word classes along-with their properties play significant role in the formation of phrases and sentences. Various part of speech classes in Sindhi include noun, pronoun, postposition, adjective, adverb, verb, conjunction, and interjection. Most of these classes are further divided into subclasses. Word classes which accept new members frequently are referred to as open word classes and classes which do not readily accept the new words are called closed word classes. Subsequent sections discuss Sindhi word classes, open and closed class types, and their characteristics.

1.1 Closed Word Classes in Sindhi

Closed word classes in a language are those classes which have relatively fixed membership. These classes usually contain small number of words as compared to open word classes. Preposition in English and postposition in Sindhi are examples of closed word class. New postpositions are rarely added. Table III-1 shows closed word classes of Sindhi. Details of these closed word classes are discussed in following sections. Different function words which do not belong to any part of speech class and have different grammatical function also belong to closed word class.

Table III-1: Closed word classes in Sindhi.

S.No.	Word Class	Examples
1.	Pronoun	tUN “you.2P.Sg”, tavhAN “you.2P.PI”, asIN “we.1P.PI”, AUN “I.1P.Sg”, uhE “they.3P.PI”, ihE “these.PI.Demon”
2.	Postposition	kHE “Dat”, tE “on.Loc”, mEN “in.Loc”, jO “of.Gen.M.Sg”, jA “of.Gen.M.PI”, jE “of.Gen.M.Sg.Obl”, mAN “from-inside.Abl”, tAN “from-upside.Abl”, TAIN “till.Loc”
3.	Conjunction	aIN “and”, yA “or”, par “but”, ta “that”, CHAkANra “because”
4.	Auxiliary, Copula and Modal Verbs	AhE “be.Cop.Sg”, tHO “Aux.Pres.M.Sg”, hA “Aux.Contra”, payO “Aux.Cont.M.Sg”, sagHE “can.Mod”
5.	Numerals, Cardinals, Ordinals, Fractals, Multipliers	hika “one”, paNja “five”, pahriUN “first”, CHahUN “sixth”, paNjUNrU “five times.M.Sg.Nom”, adHa “half”, munU “three quarters.M.Sg.Nom”, pAu “quater”
6.	Negative, Affirmative	na “no”, kona “not”, konhE “not”, hA “yes”, bHalE “yes.Affirm”
7.	Uncategorized function words	ba “also”, puNra “also”, l “also.Int”, ta “that”

1.1.1 Pronouns

Pronouns are forms that act as a kind for referring a noun and are considered closed word class. Sindhi pronouns are divided into seven different types which include demonstrative, reflexive, relative, co-relative, indefinite and wh-pronoun. Personal pronouns are further divided into first person, second person and third person pronouns. Demonstrative pronouns are further divided into proximate and remote categories. Table III-2 shows different types of pronouns in Sindhi along-with their examples.

Table III-2: Sindhi Pronouns and their subclasses.

S.No.	Type	Subtype(s)	Example
1.	Personal Pronoun	1 st Person	AUN “I.1P.Sg.Nom”, mAN “I.1P.Sg.Nom”, “we.1P.Pl.Nom”
		2 nd Person	tUN “you.2P.Sg.Nom”, tavhIN “you.2P.Pl.Nom”, “2P.Pl.Nom”
		3 rd Person	uhE “they.3P.Pl.Nom”, unhani “they.3P.Pl.Obl”
2.	Demonstrative Pronoun	Proximate	hlu “this.Demon.Prox.Sg.Nom”, ihO “it.Demon.Prox.M.Sg”
		Remote	hO “those.Demon.Rem.Pl.Nom”, uhO “it.Demon.Rem.M.Sg”
3.	Wh-Pronoun	-	kehRO “which.Wh.M.Sg.Nom”, CHA “what.Wh”
4.	Reflexive Pronoun	-	pANra “him/my self”
5.	Relative Pronoun	-	jehRO “Like.Rel.M.Sg.Nom”
6.	Co-relative Pronoun	-	tehRO “Like.Corel.M.Sg.Nom”
7.	Indefinite Pronoun	-	kO na kO “Someone.M.Sg.Nom”

First person pronouns “AUN” and “mAN” represent 1st person singular pronoun in two different dialects (vicholi/laari and utradi respectively). In the same way 2nd person pronouns “tavhIN” and “avhIN” are also used

interchangeably. Third person pronouns have ambiguity with demonstrative pronouns; for example, third person pronouns “hO” هو and “uhE” اُھي can also be demonstrative pronouns (Grierson, 1909), (Allana, 2010), however; demonstrative pronouns are always followed by a noun in a phrase or sentence and syntactically they belong to the specification of noun in LFG theory (see Section 2 and 4). Table III-2 do not show exhaustive entries, only examples of each type are given.

1.1.2 Postpositions

Another closed word class in Sindhi is postposition. Postpositions usually come after nouns, pronouns, adjectives and adverbs. They show relationship between two nouns, noun and pronoun or nouns and adjectives. There are two types of postpositions in Sindhi simple postpositions and compound postpositions. Another type of postpositions may also be considered as hidden postposition which is inflected form of noun or adverb after postpositional suffixation (represents the ablative form/case of nouns or adverbs). Table III-3 show three types of postpositions and their examples. Although Sindhi grammar writers define compound postpositions separately, but syntactically these are embedded simple postpositions.

Table III - 3: Types of Sindhi Postpositions.

S.No.	Postposition Type	Example
1.	Simple Postposition	mEN “in.Loc”, tE “on.Loc”, kHAN “from/than.Abl”, tAN “from.Loc”
3.	Compound Postposition	jE vicha mEN “in between”
4.	Hidden Postposition	GOTHAN “village.Abl”

1.1.3 Conjunction

Generally, conjunctions in Sindhi fall into copulative, concessive, adversative, conditional, interrogative, casual and final categories (Trumpp, E., 1872) but syntactically all these categories fall into two main types coordinate and subordinate conjunctions. Table III-4 show these two types along-with examples.

Table III-4: Conjunction types in Sindhi.

S.No.	Conjunction Type	Example
1.	Coordinate Conjunction	aIN “and”, yA “or”, par “but”
2.	Subordinate Conjunction	jEkaDhN “if.Desiderative”, jEkar “if”, jEtONrEk “whereas”, ta “that”, tiNhN hUNdE ba “having that”

1.1.4 Auxiliary, Copula and Modal Verbs

Auxiliary or helping verbs in Sindhi are used to mark tense and aspect. They are used with main verbs, adverbs and nouns. Examples of auxiliary verb include (“AhE”, “tHO”, “rahiyO”, “payO”). Copula verbs join noun or adjective complements with subjects; in Sindhi “AhE” is used as “be” copula verb. Modal verbs are also type of auxiliary verbs which express possibility or necessity. Examples of Sindhi modal verbs include “sagHE”, “kHapE”, and “gHurjE”. All the three types of verbs discussed above are closed class type verbs in Sindhi.

1.1.5 Uncategorized function words

These words in Sindhi are mostly used as intensifiers and belong to closed class. Examples of particles include ba, puNra, I, and ta.

1.1.6 Interjection

Interjections convey emotions in sentences. There are a small number of interjections used in Sindhi. Few examples are: vah vah, hAi hAi, afsOsi, gHORA!, aRE!, kAsh! and shAbAsh.

1.2 Open Word Classes in Sindhi

Major open classes include nouns, verbs, adjectives, and adverbs. Open Sindhi classes are discussed in following sections.

1.2.1 Nouns

Sindhi proper, common and abstract nouns belong to open class, Table III-5 show examples of three types of nouns along-with newly included words of Sindhi in these classes.

Table III-5: Types of Sindhi Nouns.

S.No.	Noun Type	Example	Newly Included Word(s)
1.	Proper Noun	Ahmed, karAcHI, pAkistAn	inTarnET, nOkiA, mAikrOsAft
		Ahmed, Karachi, Pakistan	Internet, Nokia, Microsoft
3.	Common Noun	mulka, kitAbu, shahiru	ImEI, atHAriTI, kampiUTar
		Country, book, city	Email, authority, computer
4.	Abstract Noun	aCHANra, tHadHANra	kanEkTlviTI
		Whiteness, coolness	Connectivity

1.2.2 Adjectives

Adjectives modify nouns. Various types of adjectives in Sindhi include cardinals, ordinals, multipliers, fractals and pronominal adjectives; Table III-6 show examples of such types of adjectives. Syntactically pronominal adjectives are pronouns used as noun specifiers. Syntactic agreement between adjectives and modified nouns (head noun in case of noun phrase) holds in number, gender and case. There are examples of adjectives which are adopted from other languages. For example, in the sentence “ahmed hika EnlmETEd vEbasAITa THAhl” the word “EnlmETEd” is adjective which is adopted and is transliterated from of English word “animated”. The word “KHUBaSUrat” and “sHAndAra” are also examples of adopted words in Sindhi adjectives.

Table III-6: Examples of adjectives in Sindhi.

S.No.	Adjective Type	Example(s)
1.	Cardinal	hika “one”, Ba “two”, cHAra “four”
2.	Ordinal	pihiryUN “first”, paNjUN “fifth”
3.	Multiplier	BINrU “double”, paNjUNrU “quintuple”
4.	Fractals	adHa “half”, munU “three-quarters.M.Sg.Nom”
5.	Pronominal Adjective	asIN mANrhU “we people”

1.2.3 Adverbs

Adverbs are non-subcategorized modifiers of verbs which include temporal, space, manner, negation, quantity, affirmative, noun, adjective and pronoun types. Table III-7 shows different types of adverbs in Sindhi. Space, noun and adjective adverbs can have new entries. As nouns and adjectives, themselves are open class types so is the case with noun and adjective

adverbs. For example, in the sentence “AnlAin halyO acHu” (come online) the word “AnlAin” can be considered as space adverb or noun adverb.

Table III-7: Adverb types in Sindhi.

S.No.	Adjective Type	Example(s)
1.	Temporal Adverbs	hINara “now”, pOI “then”
2.	Manner Adverbs	AhistE “slow”, DAdHO “too.M.Sg.Nom”
3.	Negation Adverb	na “no”, kOna “not”
4.	Quantity Adverb	kEtrA “how-many.M.Pl.Nom”, gHaNrA “many.M.Pl.Nom”
5.	Affirmative Adverb	hA “yes”, bHalE “yes/sure”
6.	Pronoun Adverb	hlaN “like-this”, hUNaN “like-that”
7.	Space Adverb	hitE “here”, pOyAN “behind”
8.	Adjective Adverb	“suTHO” “good”

1.2.4 Verbs and Compound Verbs

Verbs are main part of the predicate of a sentence. Main verbs are divided into intransitive and transitive types. Intransitive verbs are those verbs which are without an object in a sentence. For example, in sentence “AUN dORAN tHO” the word “dORAN” is intransitive verb. Transitive verbs are verbs with subject as well as object in a sentence. For example, in sentence “CHOkirO KHatu likHE thO” (the boy writes a letter) the word “likHE” (write) is a transitive verb as it has subject “CHOkirO” and object “kHatu”. Compound verbs in Sindhi are formed by combining nouns with verbs, adjectives/adverbs with verbs and participles with verbs. The new compound verbs can be formed by combining the nouns with verbs. Table III-8 show examples of compound verbs along-with newly formed compound verbs.

Table III-8: Sindhi Compound Verbs.

S.No.	Compound Verb	Composition	Newly adopted Compound Verb(s)
1.	rANda karaNra	Noun+Verb	cHETa-karaNra, "doing chat" ImEla-karaNra "sending email"
3.	KHUsha tHiyaNra, nArAZ karaNra	Adjective/Adverb + Verb	AnlAin-tHiyaNra "getting online", AflAin-karaNra "getting offline"
4.	kirl pavaNra, bukHiyO CHaDaNra, BudHI sagHaNra	Participle + Verb	kanEkTa-tHiyaNra "getting connected"

1.2.5 Participles

Participles are derived from verb roots. There are five different types of participles in Sindhi shown in Table III-9. Participles (excluding conjunctive participle) sometimes act as adjectives in sentences and can create ambiguity during POS tagging and syntax analysis.

Table III-9: Participles in Sindhi.

S.No.	Participle Type	Example
1.	Present Participle	dORaNdI "run.Prespart.F.Sg", DINdO "give.Prespart.Sg.M"
2.	Past Participle	paRhiyala "read.Pastpart", dORiyO "run.Pastpart.M.Sg", ladHala "find.Pastpart"
3.	Future Participle	vaNgaNrU "go.Futpart.M.Sg", viRhaNrU "fight.Futpart.M.Sg", vaTHaNriUN "take.Futpart.F.PI"
4.	Verbal Noun	pAINdaRa "one-who-puts-in", upAiNrahAra "grower", kAshlgara "designer"
5.	Conjunctive Participle	kHAI "having-eaten", paRhAI "having-taught", vaTHiO "having-got"

2. MORPHOLOGY

Sindhi is an Indo-Aryan language with rich inflectional and derivational morphology (Rahman, 2009). Sindhi morphological constructions include derivational and inflectional morphology with addition, subtraction and replacement methods. Following sections discuss Sindhi word types and their morphological construction in detail. It may be noted that word classes discussed in sections (2.3 – 2.8) are morphologically simple classes as compared to classes discussed in sections (2.9 – 2.16) where verbs, pronominal suffixes and postpositional suffixes are discussed.

2.1 Sindhi Words

Words are generally divided into two types primary or simple words and secondary words (Jatoi, 1983). Primary words (also known as minimum free forms) are not further divisible. For instance, “JANra” (knowledge) and “rastO” (path or way) are examples of primary words. Secondary words are further divided into complex and compound words. Complex words are formed by combining affixes with primary words and therefore are inflections of primary words. For example, primary word “JANra” when combined with prefix “aNra” (negation) becomes a complex word “aNraJANra” (layman). Same word when combined with suffix “U” becomes JANrU (scholar). Compound words are combinations of two or more simple words. For example, “jHaNgaBiLO” (wild cat) which is formed by combining two free form morphemes “jHaNga” (forest) and “BiLO” (cat); and “hatHakaRI” (manacle) which is formed by “hatHa” (hand)

and “kaRI” (ring) are examples of compound words. Compound words also hold the inflectional properties; however, inflections are applied on suffix words only. Native Sindhi nouns always end in a vowel (Sheikh, 2006); (exceptions are there in case of adopted words where this property may not hold). These endings not only help in identifying the gender in case of nouns but change in them can cause number inflections as well. Words can have following vowel endings. See Appendix B for roman transliteration scheme used throughout this study.

Sindhi:	ا	آ	ا	ای	اُ	او	او	اوی	اوی	
IPA:	ɑ	ɑ:	ɛ	ɛ:	ʊ	ʊ:	o	o:	ɪ	ɪ:
Roman:	a	A	i/e	I	U	U	O	aO	I	aI

2.2 Morphological Constructions in Sindhi

As discussed earlier morphology deals with word formations. For example, a simple morphological pattern in Sindhi nouns can be defined as “I → UN /ε__#” which says that words with “I” ending like in common noun (CHOkirl) are always converted into plural by replacing “I” with “UN” whenever “I” is in the end of a word; “#” and “ε” represent a word boundary and null string respectively. Resulting word will be "CHOkirUN" in this case. These patterns are then used to develop morphological rules to be used by finite state transducers for morphological modeling. Complete example is shown in Table III-10. It may be noted that this rule is for feminine singular nouns only; not a general rule for every word ending with “I” suffix.

Table III - 10: An example of FST rule.

Singular	Plural	Rule
CHOKirl	CHOkirUN	I → UN / ε ____ #.

Derivations take place when word stem is combined with a grammatical morpheme usually resulting in a different class word. For example, the adjective “bukHiyO” (hungry) is derived from noun bukHa (hunger) when suffix “yO” is added to the noun. Sindhi derivational morphology also takes place by diacritic or last vowel change. For example, nouns are derived from verbs like noun “pOkHa” (crop) is derived from verb “pOkHi” (sow) by changing of last vowel “a” to “i”.

2.3 Noun Morphology

As discussed above Sindhi nouns end in a vowel. These endings mark the gender and sometimes number of a noun. Following sections discuss noun inflections with respect to gender, number and case.

2.3.1 Gender

Nouns in Sindhi can either be masculine or feminine. This gender classification is for animate and in-animate nouns. For example, “gHaru” (house) in Sindhi is masculine and “hOtala” (hotel) is feminine. Gender of in-animate nouns is defined artificially and usually smaller things are considered feminine and larger ones are masculine (there are some exceptions can be seen in Table III-11). As discussed above gender of noun is mostly identified by last vowel/diacritic, feminine nouns mostly end with (a, A, i, and I) endings and masculine nouns usually end with (u, U, UN, O) endings; Table III-11

shows examples of masculine and feminine nouns.

Table III - 11: Sindhi Masculine and Feminine Nouns

Type	Word	Ending	Gender	English
Animate Nouns	ZAla	í (a)	F	Wife
	dunyA	í (A)	F	World
	rAti	! (i)	F	Night
	CHOkirl	ي (I)	F	Girl
	BAru	ا (u)	M	Child
	viCHUNU	ون (UNu)	M	Scorpion
	CHOkirO	او (O)	M	Boy
In-animate Nouns	Daru	ا (u)	M	Door
	darl	ي (I)	F	Window
	taNbU	ون (U)	M	Tent
	dHartl (Exception, bigger In-animate but feminine)	ي (I)	F	Earth

2.3.2 Number

Like English, Sindhi nouns also have two numbers Singular and Plural.

Number inflections depend on the gender of noun and ending vowel/diacritic.

Number inflections in feminine and masculine nouns take place differently.

Table III-12 shows examples of feminine and masculine nouns and number inflections.

Table III - 12: Number inflections in feminine and masculine nouns

Gender	Singular Noun	Plural Noun	Ending Vowel
Feminine	ZAla زال “wife.F.Sg.Nom”	ZAI-UN زالون “wives.F.PI.Nom”	ا (a)
	CHOKirl چوکری “girl.F.Sg.Nom”	CHOKir-UN چوکریون “girls.F.PI.Nom”	ا (ا) اي
	havA هوا “wind.F.Sg.Nom”	havA-UN هوائون “winds.F.PI.Nom”	ا (A)
Masculine	CHOKirO چوکرو “boy.M.Sg.Nom”	CHOKir-A چوکرا “boys.M.PI.Nom”	او (O)
	puTu پٹ “son.M.Sg.Nom”	puT-a پت “sons.M.PI.Nom”	ا (u)
	pakHI پکی “bird.M.Sg.Nom”	pakH-I پکی “birds.M.PI.Nom”	ا (ا) اي

2.3.3 Case

Sindhi grammar writers discuss different number of cases (five to eight) in Sindhi. However, corpus analysis in (Rahman, 2010) show evidence of following ten cases in Sindhi.

- (i) Nominative
- (ii) Accusative
- (iii) Dative
- (iv) Locative
- (v) Instrumental
- (vi) Participant
- (vii) Agentive
- (viii) Possessive/Genitive
- (ix) Ablative
- (x) Vocative

Nouns appear in nominative case when they act as a subject or complement of a subject in a sentence. Nouns are not inflected in nominative case and remain in their original form however number and gender inflections take place as usual. Sentence (1a) and (1b) given below show examples of

nouns in nominative case. In (1a) the noun “CHOkirO” (boy) has nominative case and in (1b) “CHOkirO” and “rANdIgara” (player) are examples of nominative case (in (1b) “rANdIgara” is complement of subject “CHOkirO”)

CHOkirO	kHAE	tHO	-----	(1a)
Boy.Nom	Eats	be.Aux.Pres		
The boy eats.				

CHOkirO	rANdIgara	AhE	-----	(1b)
Boy.Nom	player.Nom	be.Cop		
Boy is a player.				

Accusative case of noun is marked by post-position “kHE”. It may be noted that accusative case marks the direct object of the verb in a sentence. (2a) shows an example of accusative case noun where “CHOkirE” (oblique inflection of “CHOkirO”) has accusative case. Direct object can also be in nominative case as well this can be seen in (2b) where direct object CHOkirO is in nominative case. This is an example of differential object marking (DOM).

See Chapter V for detailed discussion on differential object marking.

Akbar	CHOkirE-kHE	mAriyO	-----	(2a)
Akbar-Nom	boy.Obl-Acc	Killed		
Akbar killed the boy.				

Akbar	CHOkirO	mAriyO	-----	(2b)
Akbar-Nom	boy.Nom	Killed		
Akbar killed the boy.				

In dative case noun is again marked by post-position “kHE” however, dative case marks the indirect object of a verb in a sentence. In (3) post-position “kHE” is used as a dative marker and “CHOkirE” (inflected form of CHOkirO) is an indirect object.

Akbar	CHOkirE-kHE	kitAbu	DinU	-----	(3)
Akbar-Nom	boy.Obl-Dat	book.Nom	Gave		
Akbar gave a book to the boy.					

It may be noted that post-position “kHE” acts as dative as well as accusative

case marker depending on the argument structure of verb. However, inflected form in dative and accusative case remains same if it is marked by “kHE”.

Locative case is marked by post-positions “tE”, and “mEN”, (4a) and (4b) show examples of locative case; noun in locative case is inflected like accusative and dative cases.

CHOkirE-tE kapaRO vijHa ----- (4a)
 boy.Obl-Loc cloth.Nom put.Sg
 Put a cloth on the boy.

CHOkirO kamrE-mEN AhE ----- (4b)
 boy.Nom room.Obl-Loc be.Cop.Sg
 The boy is in the room.

Instrumental case is marked by post positon “sAN”. Like in (5a) “sONTE sAN” is example of instrumental case where “sONTE” (inflected form of “sONTO”) is in instrumental case.

CHOkirO sONTE-sAN gaDahu-kHE mArE ----- (5a)
 boy.Nom cane.Obl-Inst donkey-Acc beat.Aorist
 tHO
 be.Aux.Pres
 Boy beats the donkey with cane

Participant case (sometimes called commutative case) is also formed like instrumental case with same inflection and case marker i.e. “sAN”. However, noun in participant case is animate noun where in instrumental case it is inanimate. “CHOkirE sAN” in (5b) is example of participant case.

AUN CHOkirE-sAN Ayasu ----- (5b)
 I.Nom boy.Obl-Part Came.Sg.1P
 I came with (a) boy.

Possessive/Genitive case is marked by different genitive post-positions like “jO”, “jE”, “jl” and “vArO”. Following are examples of possessive case of nouns. In these sentences “CHOkirE”, “Akbar”, and “sOna” are nouns with possessive case.

hlu CHOkirE-jO kitAbu AhE ----- (6a)
 this.Nom boy.Obl-Gen.Nom book.Nom be.Cop.Sg
 This is boy's book.

CHOkirE-jE kutE-jO paTO ----- (6b)
 boy.Obl-Gen.Obl dog.Obl-Gen.Sg.M.Nom belt.M.Sg.Nom
 Boy's dog's belt.

akbar-vArO dOstu kitHE AhE ----- (6c)
 akbar-Gen friend.Nom Where be.Cop.Sg
 where is Akbar's friend?

ihA sOna-jl ZaNjlra AhE ----- (6d)
 It.F.Sg gold-Gen Chain be.Cop.Sg
 It is a gold chain / chain of gold.

It can be observed from above examples that in accusative, dative, instrumental, participant, locative, and genitive cases nouns are inflected and their inflected form remains same (common) throughout these cases. This same inflected form in these cases is known as oblique form; moreover, in Sindhi ergativity is also reflected via oblique form of noun (See section 3.2.2 for more discussion on obliqueness and ergativity).

Ablative case expresses motion away from the source. In Sindhi ablative case in nouns is represented by inflection i.e. adding “-AN” suffix with nouns or by using “kHAN” marker with oblique noun form (Ahmed, 2007). In (7a) “gHarAN” is inflected form of “gHaru” (home) in ablative case. It may be noted that when “kHAN” marker is used; noun is inflected in oblique form as shown in (7b).

CHOkirO gHarAN AyO ----- (7a)
 boy.Nom home.Abl came
 The boy came from home.

Ahmed CHOkirE-kHAN kitAbu vartO ----- (7b)
 Ahmed boy.Obl-Abl book.Nom got.M.Sg
 Ahmed got a book from the boy.

“kHAN” marker is also used to mark agentive case. Agentive case is marked during passivization when subject becomes indirect subject (see Section 3.2).

In Sindhi case system, vocative case is formed by prefixing an interjection before nominative. For example, “aE dOsta” (O friend) and “aRE dOsta” (oh friend). Vocative is also marked by morphological change in animate nouns like in “CHOkirA!.Sg.Voc” and “CHOkiraO!.Pl.Voc”.

Table III-13 shows case marking examples of different cases of masculine common noun “CHOkirO” (boy). Ablative and vocative case inflections are also given. It may be noted that inflections during case marking are dependent on noun gender; therefore, for feminine nouns different inflections are there as given in Table III-14. Few vocative case examples are also shown in Table III-15.

2.4 Adjective Morphology

Sindhi Adjectives are declined like nouns and mostly final vowel determines the declension pattern. However, only native Sindhi adjectives are declined and adopted adjectives are not declined and remain in their original form i.e. number, gender, and case inflections do not take place for them.

Table III - 13: Noun Case Marking Inflections of Masculine Nouns

Case	Inflection Suffix		Example		
	Sg	Pl	Noun	Sg	Pl
Nominative	Ø	Ø	CHOkirO	CHOkirO	CHOkirA
Oblique	Accusative	-E	-ani	CHOkirO	CHOkir-E kHE
	Dative	-E	-ani	CHOkirO	CHOkir-E kHE
	Locative	-E	-ani	CHOkirO	CHOkir-E tE
	Instrumental	-E	-ani	sONTO	sONT-E sAN
	Participant	-E	-ani	CHOkirO	CHOkir-E sAN
	Possessive / Genitive	-E	-ani	CHOkirO	CHOkir-E jO
Ablative	-AN	-AN	gHaru	gHarAN	-
Vocative	-A	-aO	CHOkirO	CHOkirA!	CHOkir-aO

Table III - 14: Noun Case Marking Inflections of Feminine Noun CHOKirla “girl”

Case	Inflection Suffix		Example		
	Sg	Pl	Noun	Sg	Pl
Nominative	Ø	Ø	CHOKirla	CHOKirla	CHOKirUN
Oblique	Accusative	-I	-ani	CHOKirla	CHOkir-I kHE
	Dative	-I	-ani	CHOKirla	CHOkir-I kHE
	Locative	-I	-ani	CHOKirla	CHOkir-I tE
	Instrumental	-I	-ani	sONTla	sONT-I sAN
	Participant	-I	-ani	CHOKirla	CHOkir-I sAN
	Possessive / Genitive	-I	-ani	CHOKirla	CHOkir-I jO
Ablative	-AN	-AN	CHiti	CHit-AN	-
Vocative	-I	-iaO	CHOKirla	CHOkir-I!	CHOkir-iaO

Table III - 15: Sindhi vocative case examples

Num /Gender	Nominative	Vocative	English
Sg.M	BAru بار (Child)	BAra بار	O Child!
	vAdHO وادو (Carpenter)	vAdHA وادا	O Carpenter!
Sg.F	bHENru پیٹن (Sister)	bHENra پیٹن	O Sister!
Sg.M	BAra بار (Children)	BAraO بارو	O Children!
	vAdHA وادا (Carpenters)	vAdHaO وادئو	O Carpenters!
Sg.F	bHENrUN پیٹون (Sisters)	bHENraO پیٹو	O Sisters!

Table III-16 shows declension examples of native Sindhi adjective “suTHO” (good) and Table III-17 shows adopted adjective “KHUbSUrat” (beautiful) with no declension at all.

Table III - 16: Declension of Native Adjective “suTHO”

Case	Masculine	Feminine
Nom.Sg.	suTHO CHOkirO	suTHI CHOkirla
Nom.pl.	suTHA CHOkirA	suTH-iUN CHOkir-iUN
Obl.sg	suTH-E CHOkir-E	suTH-I CHOkir-I
Obl.pl	suTH-ani CHOkir-ani	suTH-iuni CHOkir-iuni
Abl.sg	suTH-E CHOkirE-kHAN	suTH-I CHOkir-I kHAN
Abl.pl	suTH-ani CHOkir-ani kHAN	suTH-iuni CHOkir-iuni kHAN
Voc.sg.	suTH-A CHOkir-A!	suTH-I CHOkir-I!
Voc.pl.	suTH-A CHOkir-aO!	suTH-iUN CHOkir-iUN!

Table III - 17: Adopted Adjective without Declension

Case	Masculine	Feminine
Nom.Sg.	KHUbSUrat CHOkirO	KHUbSUrat CHOkirla
Nom.pl.	KHUbSUrat CHOkirA	KHUbSUrat CHOkir-iUN
Obl.sg	KHUbSUrat CHOkir-E	KHUbSUrat CHOkir-I
Obl.pl	KHUbSUrat CHOkir-ani	KHUbSUrat CHOkir-iuni
Abl.sg	KHUbSUrat CHOkirE-kHAN	KHUbSUrat CHOkir-I kHAN
Abl.pl	KHUbSUrat CHOkir-ani kHAN	KHUbSUrat CHOkir-iuni kHAN
Voc.sg.	KHUbSUrat CHOkir-A!	KHUbSUrat CHOkir-I!
Voc.pl.	KHUbSUrat CHOkir-aO!!	KHUbSUrat CHOkir-iUN!

2.5 Pronoun Morphology

Pronouns are declined for number and gender and are marked by nominative case, oblique form and genitive case (genitive is separated from oblique form here due to special inflections). It may be noted that pronouns are mostly declined for number only; however, instances of gender declension can also be found (Cole, J.S., 2001). Personal pronouns are marked by number, gender and case through morphological changes. Table III-18 shows personal pronoun declension patterns. Gender declension is not there in case of first and second person pronouns in nominative case. However, third person pronouns are declined for gender in nominative and genitive cases. In oblique case gender declension is not there.

Demonstrative pronouns are marked by number, gender, and case (not marked by genitive case). Examples of different declension patterns of demonstrative pronoun are shown in Table III-19.

Table III - 18: Personal Pronoun Declensions

Number	Gender	Case/Form	1 st Person	2 nd Person	3 rd Person
Singular	Masculine	Nom	AUN “I”, mAN “I”	tUN “you”	hI “he”, ihO “this very”, uhO “that very”, hU “that/he”
		Obl	mUN “I/me”	tO “you.Sg.Obl”	hina “he.Obl”, inhE “this very.Obl”, unhE “that very.Obl” huna “that/he.Obl”
		Gen	muhiN-jO “my”, mUN-jO “my”	tuhiN-jO “your’s.M.Sg”	hina-jO “he.Obl-Gen.M.Sg”, inhE-jO “that very.Obl-Gen.M.Sg”
	Feminine	Nom	AUN “I”, mAN “I”	tUN “you”	hla “she”, ihA “this very.F”, uhA “that very.F”, hUA “that/he.F”
		Obl	mUN “I/me.Obl”	tO “you.Obl”	hina “she.Obl”, inhE “this very.Obl”, unhE “that very.Obl” huna “that/she.Obl”
		Gen	muhiNjl “my”, mUN-jl “my”	tuhiN-jl “your’s.F.Sg”	CHOkirE-jl “boy.Obl-Gen.F.Sg”
Plural	Masculine	Nom	asIN “we”	tavhIN “you.PI”, avhIN “you.PI”	hO “they”, ihE “these very”, uhE “those very”
		Obl	asAN “we.Obl”	tavhAN “you.PI.Obl”, avhAN “you.PI.Obl”	hun-ani “these.Obl”, inh-ani “these very.Obl”, unh-ani “those very.Obl”
		Gen	asAN-jO “we.Obl-Gen.M.Sg”	tavhAN-jO “you.PI.Obl-Gen.M.Sg”, avhAN-jO “you.PI.Obl-Gen.M.Sg”	CHOkirani-jO “boys.Obl-Gen.M.Sg”
	Feminine	Nom	asIN “we”	tavhIN “you.PI”, avhIN “you.PI”	hO “they”, ihE “these very”, uhE “those very”
		Obl	asAN “we.Obl”	tavhAN “you.PI.Obl”, avhAN “you.PI.Obl”	hun-ani, inh-ani, unh-ani
		Gen	asAN-jl “we.Obl-Gen.F.Sg”	tavhAN-jl “you.PI.Obl-Gen.F.Sg”, avhAN-jl “you.PI.Obl-Gen.F.Sg”	CHOkirlani-jl “girls.Obl-Gen.F.Sg”

Table III - 19: Demonstrative Pronoun Declension Patterns

		Number	Gender	Case / Form	Proximate	Remote
Singular	Masculine	Nom	hl "this.Nom", ihO "it.Prox.Nom"	hU "that.Rem.M.Sg.Nom", uhO "it.Rem.M.Sg.Nom"		
		Obl	hina "this.Obl", inhE "this very.Obl"	huna "that.Obl", unhE "that very.Obl"		
		Gen	-	-	-	
	Feminine	Nom	hla "this.F.Sg.Nom", ihA "it.F.Sg.Nom"	hUa "that.Rem.F.Sg.Nom", uhA "it.Rem.F.Sg.Nom"		
		Obl	hina "this.Obl", inhE "this very.Obl"	huna "that.Obl", unhE "that very.Obl"		
		Gen	-	-	-	
Plural	Masculine	Nom	hE "these.PI.M.Nom", ihE "it.PI.M.Nom"	hO "those.PI.M.Nom", uhE "it.Rem.PI.Nom"		
		Obl	hin-ani "these.Obl", inh-ani "it.PI.Obl"	hun-ani "those.Rem.Obl", unh-ani "those very.Obl"		
		Gen	-	-	-	
	Feminine	Nom	hl "these.F.PI", ihE "it.F.PI"	hU "those.F.PI", uhE "it.Rem.PI.Nom"		
		Obl	hin-ani "these.Obl", inh-ani "it.PI.Obl"	hun-ani "those.Rem.Obl", unh-ani "those very.Obl"		
		Gen	-	-	-	

As shown in Table III-20 some Wh-Pronouns are not inflected at all and some are declined by number, gender and case. Wh-Pronoun "kitHE" (where) is marked by ablative case only.

Table III - 20: Wh-Pronoun Declension Patterns.

S.No.	Wh-Pronoun	Morphological change(s)
1.	CHA “what”	-
2.	CHO “why”	-
3.	klaNa “how”	-
4.	kaDhaNi “when”	-
5.	kEsIN “how long”	-
6.	kitHE “where”	kitH-AN (Abl.)
7.	kehRO “which.M.Sg.Nom”	kehR-A “which.M.PI.Nom”, kehR-E “which.M.PI.Obl”, kehR-ani “which.M.PI.Obl”, kehR-I “which.F.Sg.Nom”, kehR-la “which.F.Sg.Obl”, kehR-iUN “which.F.PI.Nom”, kehRI-ani “which.F.PI.Obl”
8.	kEru “who.M.Sg.Nom”	kEra “who.M.PI.Nom”, kEra “who.F.Sg.Nom”, kEra “who.F.PI.Nom” kiNhiNh “who.Sg.Obl”, kini “who.PI.Obl”
9.	kEtrO “how-much.M.Sg.Nom”	kEtr-A “how-much.M.PI.Nom”, kEtr-E “how-much.M.Sg.Obl”, kEtr-ani “how-much.M.PI.Obl”, kEtr-I “how-much.F.Sg.Nom”, kEtr-UN “how-much.F.PI.Nom”, kEtr-la “how-much.F.Sg.Obl”, kEtr-ane “how-much.F.PI.Obl”
10.	kEDO “how-big.M.Sg.Nom”	kED-I “how-big.F.Sg.Nom”, kED-A “how-big.M.PI.Nom”, kED-UN “how-big.F.PI.Nom”, kED-E “how-big.M.Sg.Obl”, kED-ani “how-big.M.PI.Obl”, kED-la “how-big.F.Sg.Obl”, kEDI-ani “how-big.F.PI.Obl”

Reflexive pronouns are neither inflected nor marked by any case. They remain in their original form. However relative and co-relative pronouns are declined for number gender and case (nominative, oblique and sometimes ablative). Table III-21 shows different examples of relative and co-relative pronoun declension patterns.

Table III- 21: Relative and Co-Relative Pronoun Declension Patterns.

Number	Gender	Case	Relative	Co-relative
Singular	Masc.	Nom	jO “who/what.M.Sg”	sO “he/that.M.Sg”
		Obl	jiNhNi “who/what.Sg.Obl”	tiNhNi “he/that.Sg.Obl”
		Abl	jEhAN “from-who.Sg.Abl”	tEhAN “from-him.Sg.Abl”
	Fem	Nom	jA “who/what.F.Sg”	sA “he/that.F.Sg”
		Obl	jiNhNi “who/what.Obl”	tiNhNi “he/that.Obl”
		Abl	jEhAN “from-who.Sg.Abl”	tEhAN “from-him.Sg.Abl”
Plural	Masc	Nom	jE “who/what.PI”	sE “those/that.PI”
		Obl	jini “who/what.PI.Obl”	tini “those/that.PI.Obl”
		Abl	jinhAN “from-who.PI.Abl”	tinhAN “from-those.PI.Abl”
	Fem	Nom	jE “who/what.PI”	sE “those/that.PI”
		Obl	jini “who/what.PI.Obl”	tini “those/that.PI.Obl”
		Abl	jinhAN “from-who.PI.Abl”	tinhAN “from-those.PI.Abl”

Indefinite pronouns are inflected by number, gender and case. Examples of indefinite pronouns and their inflections are shown in Table III-22. Compound indefinite pronouns can also be formed by using “na” particle infixation and reduplication of indefinite pronoun like “kO-na-kO” and “kA-na-kA”. However, inflection patterns of both “kO” and “kA” remain as shown in Table III-22.

Table III- 22: Declension of Indefinite Pronouns.

S.No.	Indefinite Pronoun	Morphological change(s)
1.	kO “someone.M.Sg.Nom”	kA “someone.F.Sg.Nom”, kE “someones.PI.Nom”, kihiNhi “someone.Sg.Obl”, kini “someones.PI.Obl”
2.	filANrU “someone.M.Sg.Nom”	filANrl “someone.F.Sg.Nom”, filANrA “someones.M.PI.Nom”, filANrE “someone.M.Sg.Obl”, filANrani “someones.M.PI.Obl”, filANrla “someone.F.Sg.Obl”, filANriUN “someones.F.PI.Nom”, filANruni “someones.F.PI.Obl”

2.6 Postposition Morphology

Few postpositions are marked by number, gender and sometimes case but most remain uninflected. Examples of postposition inflections are shown in Table III-23.

Table III- 23: Postposition Morphology

S.No.	Postposition	Morphological change(s)
1.	jO “of.M.Sg.Nom”	jA “of.M.PI.Nom” jl “of.F.Sg.Nom” jUN “of.Fem.PI.Nom” jE “of.M.Sg.Obl”
2.	kHE (Nom. Sg./Pl.)	khAN “from.Abl”
3.	tE (Nom. Sg/Pl)	tAN “from-on.Abl”.
4.	mEN “in”, maNjHa “in/into”	mAN “from-in.Abl”, maNjHAN “from-in.Abl”
5.	vaTa “beside”	vaTAN “from-beside.Abl”

2.7 Adverb Morphology

Generally, adverbs are not inflected (Anderson, 1985), (Alexiadou, 1997), (Evans, 2000); however, in Sindhi adjectives sometimes behave like adverbs; for example, the word suTHO (good) is used as adjective in sentence “suTHO CHOkiR O dORE thO” (Good boy runs) but just by changing the order of words in sentence “CHOkiR O suTHO dORE thO” (The boys runs well) it becomes adverb. Adjectives when used as adverbs hold the morphological properties of adjectives (interestingly syntactic agreement of adjectives also holds). Manner and quantity adverbs are mostly adjectives acting as adverbs and are marked by number, gender, and case (Butt, M., et al. 2016). Place adverbs are also marked by ablative case. Noun adverbs can be considered as place adverbs and hold the morphological properties of nouns. The pronoun

adverbs can also be considered as a subset of manner adverbs and do not change their form however, few relative and co-relative pronouns when used as adverbs can have gender and number declensions. For example jihRO “as.Rel.M.Sg.Nom” and its correlative counterpart tihRO “as.Correl.M.Sg.Nom” can have number, gender and case inflections like manner adverbs. In the same way quantity adverbs can also have relative and correlative forms. For example jEtrO “as-much.Rel.M.Sg.Nom” and tEtrO “that-much.Correl.M.Sg.Nom” can also have number, gender and case declensions.

Adverb inflections are shown in Table III-24.

Table III- 24: Adverb Inflections

S.No.	Adverb Type	Example(s)	Case Inflection
1.	Place Adverbs	hitE “here”, hETHa “below/down”	hitAN “from-here.Abl”, hETHAN “from-down.Abl”
2.	Temporal Adverbs	kaDahiN “when”, hINara “now”	-
3.	Manner Adverbs	cHaNgU “better”, suTHO “good”	Morphological properties of adjectives
4.	Negation Adverb	na “no”, kOna “not”	-
5.	Quantity Adverb	keEtrO “how-much”, gHaNU “much/more”	Morphological properties of adjectives
6.	Affirmative Adverb	hA “yes”, bHaIE “yes/sure”	-
7.	Pronoun Adverb	hlaN “like-this”, hUNa “like-that”	-
8.	Noun Adverb	aNdaru “inside”	Morphological properties of Nouns

2.8 Conjunction and Interjection Morphology

Conjunction is closed class in Sindhi and both types coordinate and subordinate conjunctions are neither inflected nor marked by any case. Interjections are also not marked or inflected, they are also closed class words and may have irregular inflections.

2.9 Sindhi Verb Morphology

Sindhi has complex verbal constructions; Sindhi verbs are marked by aspect, tense, mood, gender and number. Root verb is always in second person imperative singular form for example “likHu” (write) (some linguists consider infinitive forms of verbs as roots but as we will see in following sections that infinitive is derived from second person singular imperative). Sindhi verb forms include participle, main verb, auxiliary and copula verb constructions (Cole, J. S., 2001). Verbal components may appear in different combinations determined by tense aspect and mood.

Following sections discuss formation of imperatives, infinitives, participles, transitive verbs, causative verbs, tense, aspect, mood, and pronominal suffixes in detail. Since verb roots are 1st person imperative form the discussion starts with imperatives.

2.9.1 Imperatives

Imperatives in Sindhi can be divided into following five categories:

- i) Second person imperative form

- ii) Polite imperative
- iii) Desiderative
- iv) Prohibitive
- v) Requestive polite imperative

Second person imperative form is considered as a root; polite and desiderative forms are morphologically inflected forms of root. Last two imperative types are syntactically formed by placing either negation before any of the first three imperative forms or intensifiers “ba”, “na”, “ta” after the imperative. It may be noted that “na” is a negation if used before a verb and an intensifier if used immediately after verb. For example, prohibitive is formed as “na likHu” (don’t write) see example (8a) and requestive polite is formed by placing “na” following the imperative form like “likHu=na” (please write); though this form needs further investigation but it seems that in this case “na” is acting as an intensifier clitic. Same is the case with other intensifiers “ta” and “ba” see examples 8b to 8d.

Na	likhu	-----	(8a)
Not	write.Imp.2P.Sg		
Don’t write.			
likhu=na		-----	(8b)
write.2P.Sg.Imp=Int			
Please write.			
likhu=ta		-----	(8c)
write.2P.Sg.Imp=Int			
Please write.			
likhu=ba		-----	(8d)
write.2P.Sg.Imp=Int			
Please write.			

First three imperative forms are discussed in following sub sections.

2.9.2 Second Person Imperative (The Root)

Generally, root is second person singular form of verb. For example, "likHu", which is further inflected to form various morphological patterns of verb including other types of imperatives. Some linguists distinguish between the root and second person imperative form by inflecting the root with diacritic; like "likH" to "likHu". However, the verb with last diacritic which is the second person imperative form can be considered as root and henceforth this will be referred to as root. Table III-25 shows different examples of second person imperative verb forms/roots. These verbs are categorized into different morphological classes/categories as each category has different conjugation pattern.

Table III - 25: Verb Roots

Category	Root	Translation
1	likHu	Write
2	mAri	Kill / beat
3	kHAu	Eat
4	Kari	Do
5	DE	Give

2.9.3 Polite Imperative

Polite imperative is formed by inflecting the root (i.e. 2P.Sg.Imp) and replacing the last diacritic with polite imperative marker "-iji". Table III-26 shows polite imperative formation examples of different morphological classes of root. This imperative form is also known as future form as the work intended to be carried out in (near) future.

Table III - 26: Polite and more-polite imperative formation

Category	Root	Polite	Suffix Change
1.	likHu	likH-iji	u → iji
2.	mAri	mAr-iji	i → iji
3.	kHAu	kHA-iji	u → iji
4.	Kari	ka-ji	ri → iji
5.	DE	D-iji	E → iji

2.9.4 Desiderative

Desiderative imperative form shows the expression of desire along-with command. This is formed by adding the suffix “-ijANI” to root. Table III-27 shows different examples of passive desiderative formation.

Table III-27: Desiderative formation

Categor	Root	Desiderative	Suffix Change
1.	likHu	likH-ijANI	u → ijANI
2.	mAri	mAr-ijANI	i → ijANI
3.	kHAu	kHA-ijANI	u → ijANI
4.	Kari	ka-jANI	ri → jANI
5.	DE	D-ijANI	E → ijANI

All imperatives are marked by number with same inflection rule i.e. “u → O” for “u” ending; and “[i,E] → iyO” in case of “i” or “E” ending. Gender is not marked in case of imperatives. Table III-28 shows different examples of number marking on imperatives.

Table III-28: Number marking on imperatives.

Attributes	Root “likHu”	Root “mAri”	Root “DE”
Masc/Fem, Sg	likHu	mAri	DE
Masc/Fem, Pl	likH-O	mAr-iyO	D-iyO

2.10 Formation of Infinitive

Some linguists consider infinitives as root verbs in Sindhi but these infinitives are derived from 2nd person imperatives by applying simple morphological rules. Verb roots or stems (stems in case of passives) are inflected and suffix “-aNru” or “-iNru” is added depending on the last diacritic “u” or “i” respectively. Table III-29 shows various examples of infinitive formation along-with formation rules. Infinitives are not marked by gender or number.

Table III-29: Formation of infinitives

Category	Root	Infinitive	Suffix Change
1.	likHu	likH-aNru (to write)	u → aNru
2.	mAri	mAr-aNru (to kill)	i → aNru
3.	kHAu	kHA-iNru (to eat)	u → iNru
4.	Kari	Kar-aNru (to do)	i → aNru
5.	DE	D-iyaNru (to give)	E → iyaNru

2.11 Formation of Participles

Participles are derived from verb roots. There are five different types of participles in Sindhi namely present participle, past participle, future participle, verbal noun and conjunctive participle. Following sections discuss formation of these participles in detail.

2.11.1 Present participle

Present participles are formed by “-aNd-” or “-INd-” infixation depending on ending vowel of verb root followed by gender and number inflection suffix. Table III-30 given below shows different examples of present participle

formation from verb roots and their formation rules.

Table III-30: Formation of Present Participle

Root	Number	Gender	Present Participle	Suffix Change
mAri (Kill)	Masc	Sg	mAr-INd-O	i → INd-O
		Pl	mAr-INd-A	i → INd-A
	Fem	Sg	mAr-INd-I	i → INd-I
		Pl	mAr-INd-iUN	i → INd-iUN
likHu (Write)	Masc	Sg	likH-aNd-O	u → aNd-O
		Pl	likH-aNd-A	u → aNd-A
	Fem	Sg	likH-aNd-I	u → aNd-I
		Pl	likH-aNd-iUN	u → aNd-iUN
DE (Give)	Masc	Sg	D-INd-O	E → INd-O
		Pl	D-INd-A	E → INd-A
	Fem	Sg	D-INd-I	E → INd-I
		Pl	D-INd-iUN	E → INd-iUN

2.11.2 Past participle

Past participle is a form of verb indicates the past or completed action or event (also referred as perfective participle) (Cole, J. S., 2001). Past participles in Sindhi can have four different forms. Root verbs with “-u” or “-i” endings are converted to past participles by “-yO” suffixation; for example, past participle “likH-yO” (written/wrote) can be formed from the root verb “likHu” (write). Other form of past participle is formed by “-yala” suffixation for example “likHu” to “likH-yala” (written). Third form of past participle is formed by internal phonetic change (infixation) along-with suffixation “-O” or “-ala”; like in case of “Disa” (see) to “Di-TH-O” (seen/saw) and “Di-TH-ala” (seen). These roots/stems can be grouped into various categories which define different internal changes for past-participle formations. Last form of past participle is inflected form of verb root in which last consonant along-with its diacritic is

deleted and then past participle suffix (“-yO”, “-yala”, or “-iyala”) is added. For example, “vaNga” (go) to “va-yala” (gone). Past participles act as adjectives if they occur before noun in sentences. Apart from “-yala” and “-iyala” ending participles all other past-participle forms are marked by number and gender like adjectives and nouns (syntactically they must agree with respective clausal elements). Various past-participle formation patterns are shown in Table III-31.

Table III - 31: Formation of Past Participle.

Root	Past Participle	
mAri	mAri-yO / -I -M.Sg /-F.Sg	mAri-yA / -yUN -M.PI /-F.PI
	mAri-yala	-
likHu	likH-iyO / likH-I -M.Sg/-F.Sg	likH-iyA /-iyUN -M.PI/-F.PI
	likH-iyala	-
DE	D-in-U / -I -M.Sg/-F.Sg	D-in-A/-iyUN -M.PI/-F.PI
	D-in-ala	-
Disa	Di-TH-O /-I -M.Sg/-F.Sg	Di-TH-A / -iyUN -M.PI/-F.PI
	Di-TH-ala	-
vaNga	va-yO / -I -M.Sg/-F.Sg	va-yA / -yUN -M.PI/-F.PI
	va-yala	-

2.11.3 Future participle

Future participle in Sindhi is formed by aNrU (۽) suffixation in verb root; like “likHu” to “likH-aNrU”. Future participle is also act as an adjective when used before noun in a sentence. For example, in sentence “hla DHaGO mAr-aNrU AhE” (this bull is to be killed) the “mAraNrU” is derived future participle from “mAri” (kill); and in sentence “hla mAr-aNrU DHaGO AhE” (this is a killer bull) mAr-iNU is an adjective. Table III - 32 shows different future

participle formation patterns.

Table III - 32: Formation of Future Participle

Root	Future Participle	Formation Rule
	(-M.Sg / -F.Sg)	(-M.PI / -F.PI)
likHu “write.2P.Imp”	likH-aNr-U / -I	likH-aNr-A / -iyUN
mAri “kill/beat.2P.Imp”	mAr-aNr-U / -I	mAr-aNr-A / -iyUN
kHAu “eat.2P.Imp”	kHA-iNr-U / -I	kHA-iNr-A / -iyUN
Kari “do.2P.Imp”	Kar-Nr-U / -I	Kar-Nr-A / -iyUN
DE “give.2P.Imp”	D-iyaNr-U / -I	D-iyaNr-A / -iyUN

2.11.4 Verbal Noun

Verbal nouns are derived from verbs and act as nouns in a sentence.

Following are different ways to derive verbal nouns from verbs.

- a. Verbs with “-u” ending are converted into verbal nouns by adding the suffix “-andaRu” for example verb “likHu” (write.2P.Imp.Sg) to likHandaRu (writer.VbN).
- b. Verbs with “-i” ending are converted to verbal nouns by adding “-INdaRu” suffix; for example, verb “mAri” (kill.2P.Imp.Sg) to mAriNDaRu (killer/beater.VbN).
- c. Infinitives are converted into verbal noun by adding “-hAru” suffix; for example, “halaNra” (to walk) to “halaNrhhAru” (one who walks).
- d. Verbs are also converted into verbal nouns by adding the vArO clitic following the infinitive; for example, halaNra (walk.Inf) to halaNra=vArO (one who walks).

These derivations act as adjectives when used before nouns in sentences. Table III - 33 shows different examples of verbal noun formations.

Table III - 33: Formation of Verbal Noun

Root/Stem	Derived Verbal Noun
likHu	likH-andaRu (writer)
mAri	mAr-INDaRu (killer)
halaNra	halaNra-hAru (walker)
kHAiNra	kHAiNra-vArO (eater)

2.11.5 Conjunctive Participle

Conjunctive participles combine two separate clauses into a single sentence to show a successive sequence of two events (Schmidt, R. L., 1999). There are three different forms of conjunctive participles in Sindhi. These three participle formations are given in Table III-34. Suffix “-I” is used with main verbs, suffix “-E” is used with causative forms of verbs and suffix “-iON” is also used with some main verbs. Examples of sentences with these types of conjunctive participles are given below (9a – 9c).

CHOkirO mAnl kHA-I iskUla vayO ----- (9a)
 Boy.Nom Meal.Nom eat-ConjPart School went.M.Sg
 The boy went to school after eating meal.

ustAda Bara paRhA-rA-E gHaru vayO ----- (9b)
 Teacher.Nom Kids Teach-caus-ConjPart home.Nom.M.Sg went.M.Sg
 Teacher went home after causing kids learn.

nOkaru bHAJI vaTH-iON gHaru AyO ----- (9c)
 Servant.Nom Vegetable.Nom buy/take-ConjPart home.Nom came.M.Sg
 Servant came home after (buying / taking /having) vegetable.

It may be noted that “-iON” suffix can replace other two suffixes i.e. kHA-I “eat-ConjPart” and kHA-iON “eat-ConjPart” are valid conjunctive participles, so is the case with kHA-rA and vaTHu as shown in Table III-34.

Table III - 34: Conjunctive Participle Formation

Root / stem	Conjunctive Participle	Suffix Change	Alternate Construction
kHAu “eat.2P.Imp.Sg”	kHA-I “eat-ConjPart”	u → I	kHA-iON “eat-ConjPart”
kHA-rA “eat.2P.Imp.Caus”	kHA-rA-E “eat-Caus-ConjPart”	ε → E	kHA-rA-iON “eat-Caus-ConjPart”
vaTHu “buy/take.2P.Imp.Sg”	vaTH-iON “buy/take-ConjPart”	u → iON	vaTH-I “buy/take-ConjPart”

2.12 Transitive and Causative Formations

Intransitive verb roots are inflected to form transitive verbs. For example, intransitive verb forms with starting diacritic a (í) are inflected to transitive verbs by in-fixation of diacritic “a” (í) to “A” (í). Various such examples of transitive verb formation are shown in Table IV-26. Causative stems in Sindhi are formed by root inflections like in Urdu and Hindi (Haq, A., 1991); (Bhatt & Embick 2003), (Butt, King, & Maxwell., 2003). Causative forms are considered valence increasing forms and formation of transitive from intransitive verbs discussed above is an example of direct-causative formation. Sindhi grammar writers (Allana, G. A., 2010) discuss interesting and complicated causative patterns which can have mono-transitive, bi-transitive, tri-transitive, or even multi-transitive forms. For example, the intransitive verb “mari” (die) can be changed to mono-transitive (direct causative) form to “mAri” (kill) by the formation rule of category 1 in Table III-35 above. The stem “m-A-ri” is further inflected to di-transitive form mA-r-A-i (cause killing by someone). In this formation r-A morpheme is repeated for further inflections and forms tri-transitive mA-rA-rA-i and quadri-transitive mA-rA-rA-rA-i formations. Following examples demonstrate these causative constructions.

Table III - 35: Intransitive to transitive formation

Category	Intransitive	Transitive	Suffixation / infixation
1	mari “die.2P.Imp.Sg”	m-A-ri “kill.2P.Imp”	a → A
2	pHiri “spin.2P.Imp.Sg”	pH-E-ri “turn.2P.Imp”	i → E
3	bHuri “get-in-pieces.2P.Imp.Sg”	bH-O-ri “make-in-pieces.2P.Imp”	u → O
5	chaO “say.2P.Imp.Sg”	cha-vAi “say.2P.Caus.Sg”	O → vAi
6	utHa “awake.2P.Imp.Sg”	utH-Ari “awake.2P.Caus.Sg”	a → Ari

CHOkirE gADI halAI ----- (10a)
 boy.Obl car.F.Sg.Nom drive.PastPart.F.Sg
 The boy drove the car.

CHOkirE gADI halArAI ----- (10b)
 boy.Obl car.F.Sg.Nom drive.Caus.F.Sg
 The boy caused to drove the car (by someone).

CHOkirE gADI halArArAI ----- (10c)
 boy.Obl car.F.Sg.Nom drive.Caus2.F.Sg
 The boy caused to drove the car (by someone who actually caused it by someone else).

The increase of arguments is made complicated when transitivity refers more and more arguments. These constructions need further investigation. Although the grammar writers discuss tri-transitive and more patterns but the corpus analysis (Rahman, 2010) doesn't show any evidence of such patterns which have tri-transitive or more constructions.

2.13 Tense and Aspect

Comrie (1985) defines tense as grammaticalization of location in time.

In other words, it refers to the time of the event in relation to speech and reference time. Present, past and future are three tenses which refer the actual location in time. In Sindhi, tense is marked by main verb inflections, auxiliary verbs (tense markers) or combination of both.

Aspect refers to internal temporal structure of the situation which describes that whether the action is successfully completed (perfective aspect) or not (imperfective aspect). According to J.S. Cole (2001) aspect in Sindhi can be divided into three major categories namely unspecified, perfective, and imperfective (imperfective-habitual, imperfective-continuous). Hence the aspect defines the completeness, continuity, habituality and iterativeness of the situation. Various aspectual markers are used to mark the aspect in Sindhi.

Following sections discuss basic tenses along-with examples of perfective and imperfective constructions. The discussion starts with aorist formation as aorist provide base for various tense formations.

2.13.1 Aorist (Old Present) formation with Unspecified Aspect

In Sindhi aorist (also referred as old-present) form of verb is formed by suffixation in verb roots. Aorist plays important role in tense formation and provides basis for various formations of present tense and contrafactual mood. Suffixation depends upon last vowel/diacritic of verb root/stem. Verb root combined with suffix forms the base-form with unspecified aspect. Table III-36 shows different formations of aorist.

Table III - 36: Formation of Aorist

Category	Root/Stem	Aorist	Pronoun Type	Suffix Change
1	likHu “write.2P.Imp.Sg”	likH-AN	1P.Sg.	u → AN
		likH-UN	1P.Pl.	u → UN
		likH-IN	2P.Sg.	u → IN
		likH-O	2P.Pl.	u → O
		likH-E	3P.Sg.	u → E
		likH-ani	3P.Pl.	u → ani
2	mAri “kill.2P.Imp.Sg”	mAr-iyAN	1P.Sg.	i → iyAN
		mAr-iyUN	1P.Pl.	i → iyUN
		mAr-IN	2P.Sg.	i → IN
		mAr-iyO	2P.Pl.	i → iyO
		mAr-E	3P.Sg.	i → E
		mAr-iyani	3P.Pl.	i → iyani
3 (Aux)	AhE “is/am/be.Pres”	Ah-iyAN	1P.Sg.	E → iyAN
		Ah-iyUN	1P.Pl.	E → iyUN
		Ah-IN	2P.Sg.	E → IN
		Ah-iyO	2P.Pl.	E → iyO
		Ah-E	3P.Sg.	E → E
		Ah-ini	3P.Pl.	E → ini

2.13.2 Present Tense

Present tense in Sindhi is marked by auxiliary verb “tHO” following the main verb in aorist (old present) form. The aorist form is not further inflected however; auxiliary verbs are further inflected and marked by gender and number. Table IV-28 shows conjugation of present tense auxiliaries. Examples of present tense sentences using aorist forms of Table III-36 and auxiliary forms of Table III-37 are given in Table III-38. These examples only show masculine forms of present tense auxiliary. For feminine constructions, present tense auxiliary will be replaced by feminine inflected forms and rest of the sentence constructions will remain same as aorist forms are not marked for

gender. It can be seen in Table III-38 that aorist agrees in person and number while the present tense auxiliary thO only agrees in number and gender.

Table III-37: Present Tense Auxiliary Conjugation

Present Tense Aux	Pronoun Type	Masculine	Feminine
tHO “be.Pres.M.Sg”	1P.Sg.	tH-O	tH-I
	1P.PI.	tH-A	tH-iyUN
	2P.Sg.	tH-O	tH-I
	2P.PI.	tH-A	tH-iyUN
	3P.Sg.	tH-O	tH-I
	3P.PI.	tH-A	tH-iyUN

Table III - 38: Present Tense Formation Examples

1.	AUN	likH-AN	tHO
	I	Write.1P.Sg.Aorist	Masc.Sg.Pres.Aux
	I write.		
2.	asIN	likH-UN	tHA
	We	Write.1P.PI.Aorist	Masc.PI.Pres.Aux
	We write	.	
3.	tUN	likH-IN	tHO
	You.Sg	Write.2P.Sg.Aorist	Masc.Sg.Pres.Aux
	You write.		
4.	tvahIN	likH-O	tHA
	You.PI	Write.2P.PI.Aorist	Masc.PI.Pres.Aux
	You write.		
5.	hU	likH-E	tHO
	He	Write.3P.Sg.Aorist	Masc.Sg.Pres.Aux
	He writes.		
6.	uhE	likH-ani	tHA
	They	Write.3P.PI.Aorist	Masc.PI.Pres.Aux
	They write.		

2.13.3 Past Tense

Past tense in Sindhi is marked by past participle form of verb; followed by past tense auxiliary. For example, “mUN kHatu likH-iyO huO” (I wrote a

letter) in which likH-iyO is past participle form of verb likHu (write) and “huO” (was) is past tense form of auxiliary “AhE” (is). It may also be noted that verb can either be transitive or intransitive which can cause different morphological changes in verb itself and subject (SUBJ) (pronouns may be replaced by their oblique forms) in the sentence. It may also be noted that aorist form is not used in past tense. Also, the verb is marked by number and gender in past tense. Table III-39 show examples of transitive verbs in past tense, and morphological changes caused by them. Examples are shown with singular masculine form of likHu (i.e. likH-iyO); plural masculine form will have “-iyA” inflection instead of “-iyO” and the sentence will look like “mUN kHata likH-iyA huA” (I wrote letters). Gender and number inflections of verb likHu are shown in Table III-40.

Table III - 39: Transitive verb “likHu” with past indefinite tense

Pron Type	Sentence with transitive verb likH-iyO (Sg.Masc)	Nom → Obl
1 st PSg	mUN kHatu likH-iyO huO (I wrote a letter)	AUN → Mun
1 st PPI	asAN kHatu likH-iyO huO (we wrote a letter)	asIN → asAN
2 nd PSg	tO kHatu likH-iyO huO (you wrote a letter)	tUN → To
2 nd PPI	tavhAN kHatu likH-iyO huO (you wrote a letter)	tavhIN → tavhAN
3PSg	huni kHatu likH-iyO huO (he wrote a letter)	hU → huni
3PPI	unhani kHatu likH-iyO huO (they wrote a letter)	uhE → unhani

Table III - 40: Number and Gender Suffixes for Past Tense Verbs

	Masc.Sg.	Masc.Pl.	Fem.Sg.	Fem.Pl.
Prefix	-iyO	-iyA	-I	-iyUN
Full form	likH-iyO	likH-iyA	likH-I	likH-iyUN

Intransitive verbs are conjugated in past tense and inflections reflect the number, gender, and person. Intransitive suffixes are shown in Table III-41.

Table III-42 shows conjugation examples of intransitive verb “dORi”. It may be noted that when subject is in oblique form transitive verbs are used in their past-participle form and only number inflection takes place with singular “-yO” and plural “-yA” morphemes.

Table III - 41: Past Tense Suffixes for Intransitive Verbs

Person & Number	Masc.	Fem.
1P.Sg.	-ya-si	-ya-si
1P.Pl.	-yA-sIN	-yUN-sIN
2p.Sg.	-EN	-IEN
2P.Pl.	-yA	-yUN
3P.Sg.	-yO	-I
3P.Pl.	-yA	-yUN

2.13.4 Future Tense

Base of future tense is present participle form which changes its form according to number, gender and person. Conjugation takes place in transitive, intransitive and auxiliary verbs. Table III-43 shows future tense suffixes. Different examples of transitive, intransitive and auxiliary verbs in future tense form are shown in Table III-44, III-45 and III-46 respectively. Auxiliary “AhE” has equivalent “hUNdO” auxiliary in future tense.

Table III - 42: Past Tense Conjugation Examples of Intransitive Verb dOri

Masc		Fem	
Aun	dORi-ya-si I run-past.Sg-1P.M.Sg	aUN	dORi-ya-si I run-past.Sg-1P.F.Sg
I ran.		I ran.	
AsIN	dORi-yA-sIN We run-Past.PI-1P.M.PI	asIN	dORi-yUN-sIN We run-Past.PI-1P.F.PI
We ran.		We ran.	
tUN	dOri-EN You.Sg run-Past.Sg.2P.M	tUN	dOri-IEN You.Sg run-Past.Sg.2P.F
You ran.		You ran.	
tavhIN	dOri-yA You.PI run-Past.PI.2P.M	tavhIN	dOri-yUN You.PI run-Past.PI.2P.F
You ran.		You ran.	
hU	dORi-yO He run-Past.Sg.3P.M	hUa	dORi-I She run-Past.Sg.3P.F
He ran.		She ran.	
uhE	dORi-yA They run-Past.PI.3P.M	uhE	dORi-yUN They run-Past.PI.3P.F
They ran.		They ran.	

Table III - 43: Future Tense Suffixes

Person & Number	Masc.	Fem.
1P.Sg.	-aNd-a-si -Fut-M-1P.Sg	-aNd-la-si -Fut-F-1P.Sg
1P.PI.	-aNd-A-sIN -Fut-M-1P.PI	-aNd-iyUN-sIN -Fut-F-1P.PI
2p.Sg.	-aNd-EN -Fut-M.2P.Sg	-aNd-IEN -Fut-F.2P.Sg
2P.PI.	-aNd-A -Fut-M.2P.PI	-aNd-iyUN -Fut-F.2P.PI
3P.Sg.	-aNd-O -Fut-M.3P.Sg	-aNd-I -Fut-F.3P.Sg
3P.PI.	-aNd-A -Fut-M.3P.PI	-aNd-iyUN -Fut-F.3P.PI

Table III - 44: Future Tense Morphology of Transitive Verb “likHu”

Transitive Verb	Future Masc.	Future Fem.
likHu Write.2P.Sg	likH-aNd-a-si -Fut-M-1P.Sg	likH-aNd-ia-si -Fut-F-1P.Sg
	likH-aNd-A-sIN -Fut-M-1P.PI	likH-aNd-iyUN-sIN -Fut-F-1P.PI
	likH -aNd-EN -Fut-M.2P.Sg	likH -aNd-IEN -Fut-F.2P.Sg
	likH -aNd-A -Fut-M.2P.PI	likH -aNd-iyUN -Fut-F.2P.PI
	likH -aNd-O -Fut-M.3P.Sg	likH -aNd-I -Fut-F.3P.Sg
	likH -aNd-A -Fut-M.3P.PI	likH -aNd-iyUN -Fut-F.3P.PI

Table III - 45: Future Tense Morphology of Intransitive Verb “dORi”

Intransitive Verb	Future Masc.	Future Fem.
dORi run.2P.Sg	dOR-aNd-a-si -Fut-M-1P.Sg	dOR-aNd-ia-si -Fut-F-1P.Sg
	dOR-aNd-A-sIN -Fut-M-1P.PI	dOR-aNd-iyUN-sIN -Fut-F-1P.PI
	dOR-aNd-EN -Fut-M.2P.Sg	dOR-aNd-IEN -Fut-F.2P.Sg
	dOR-aNd-A -Fut-M.2P.PI	dOR-aNd-iyUN -Fut-F.2P.PI
	dOR-aNd-O -Fut-M.3P.Sg	dOR-aNd-I -Fut-F.3P.Sg
	dOR-aNd-A -Fut-M.3P.PI	dOR-aNd-iyUN -Fut-F.3P.PI

Table III - 46: Future Tense Morphology of Auxiliary “AhE”

Auxiliary Verb	Future Masc.	Future Fem.
AhE “is/be.Cop.Pres”	hU-Nd-a-si -Fut-M-1P.Sg	hU-Nd-la-si -Fut-F-1P.Sg
	hU-Nd-A-sIN -Fut-M-1P.PI	hU-Nd-iyUN-sIN -Fut-F-1P.PI
	hU-Nd-EN -Fut-M.2P.Sg	hU-Nd-IEN -Fut-F.2P.Sg
	hU-Nd-A -Fut-M.2P.PI	dOR-aNd-iyUN -Fut-F.2P.PI
	hU-Nd-O -Fut-M.3P.Sg	dOR-aNd-i -Fut-F.3P.Sg
	hU-Nd-A -Fut-M.3P.PI	dOR-aNd-iyUN -Fut-F.3P.PI

2.13.5 Perfective Aspect

Base form of perfective aspect is perfective-past-participle with “-iyO” ending. The perfective construction “likH-iyO” (wrote) is therefore an example of perfective aspect with unspecified tense. Table III-47 shows various aspectual forms of verb “likHu” in present, past and future tenses. (These aspects of verb are not explicitly defined as aspect by conventional grammar writers of Sindhi but are considered tense subtypes). Perfective aspect in present tense is marked by perfective-past-participle followed by present tense auxiliary “AhE” (in case of transitive verbs) and aorist form of “AhE” (in case of intransitive verbs). In the same way perfective aspect in past and future tenses is formed, the only difference is the tense auxiliary used. Conventional grammar writers of Sindhi refer present-perfective as immediate past tense, past-perfective as remote past tense and future perfective as past dubitative tense.

Table III - 47: Perfective Aspect Patterns in Different Tenses by using Perfective Participle

Tense	Verb Type	Base Form	Example
Unspecified	Transitive	Perfective-past-participle	likHi-yO write.Perf
	Intransitive		dORi-yO run.Perf
Present	Transitive	perfective-past-participle + present tense Aux	likHi-yO AhE Write.Perf.M.Sg is.Pres
	Intransitive	perfective-past-participle + old-present Aux	dORi-yO AhyAN run.Perf.M.Sg is.OldPres
Past	Transitive	perfective-past-participle + past tense Aux	likHi-yO huO write.Perf.M.Sg was.Past.M.Sg
	Intransitive	perfective-past-participle + past present Aux	dORiyO huO run.Perf.M.Sg was.Past.M.Sg
Future	Transitive	perfective-past-participle + Future tense Aux	likHi-yO hUND write.Perf.M.Sg is.Fut.M.Sg
	Intransitive	perfective-past-participle + Future tense Aux	dORiyO hUND run.Perf.M.Sg is.Past.M.Sg

Table III-47 shows perfective aspect patterns in different tenses. Number and gender agreement holds between main and auxiliary verbs in case of past and future perfective constructions. Different example sentences of perfective constructions with transitive verb likHu are shown in (11a) to (11d).

mUN kitAbu likHi-yO ----- (11a)
I.Obl book.M.Sg.Nom write-Perf.M.Sg
I written a book. (Unspecified tense)

mUN kitAbu likHi-yO AhE ----- (11b)
I.Obl book.M.Sg.Nom write-Perf.M.Sg is.Aux.Pres

I have written a book.

mUN	kitAbu	likHi-yO	huO -----	(11c)
I.Obl	book.M.Sg.Nom	write-Perf.M.Sg	was.Aux.Past	
I had written a book.				

mUN	kitAbu	likHi-yO	hUND _O -----	(11d)
I.Obl	book.M.Sg.Nom	write-Perf.M.Sg	is.Aux.Fut	
I will have written a book.				

Another interesting construction of perfective aspect in Sindhi is by using compound-verb in which a conjunctive-participle is followed by perfective-past-participle auxiliary (used as completion marker). In this formation, main verb is changed to conjunctive participle form and is not marked by number or gender; however, the following completion marker may be marked by number and gender depending on different subject / object cases (case here refers grammatical case). For example, in present-perfective construction “likH-iyO AhE” the equivalent present-perfective construction is “likHI vartO AhE”. This construction can be simply extended to past and future tenses by just replacing the auxiliary tense marker. Table III-48 shows such perfective constructions. Example sentences are shown in (12a – 12c).

CHOkirE	kitAbu	likHI-vartO	AhE -----	(12a)
boy.Obl	book.M.Sg.Nom	write.ConjPart-CompMark.M.Sg	is.Aux.Pres	
Boy has written a book.				

CHOkirE	rasId	likHI-vartI	hul -----	(12b)
boy.Obl	reciept.F.Sg.Nom	write.ConjPart-CompMark.F.Sg	was.Aux.Past.F.Sg	
Boy had written a receipt.				

Table III - 48: Perfective Aspect Patterns with Conjunctive Participle and Completion Marker Auxiliary

Tense	Verb Type	Base Form	Example
Present	Transitive		likHI-vartO-AhE write.ConjPart-got.Aux.Perf-is.Aux.Pres
	Intransitive		dORI-vartO-AhE run.ConjPart-got.Aux.Perf-is.Aux.Pres
Past	Transitive	Conjunctive-participle+Perfective-past-participle-Aux + Tense Auxiliary	likHI-vartO-huO write.ConjPart-got.Aux.Perf-was.Aux.Past
	Intransitive		dORI-vartO-huO run.ConjPart-got.Aux.Perf-was.Aux.Past
Future	Transitive		likHI-vartO-hUND0 write.ConjPart-got.Aux.Perf-is.Aux.Fut
	Intransitive		dORI-vartO-hUND0 run.ConjPart-got.Aux.Perf-is.Aux.Fut

CHOkirla kitAbu likHI-vartO hUND0 ----- (12c)
 girl.Obl book.M.Sg.Nom write.ConjPart-CompMark.M.Sg is.Aux.Fut.M.Sg
 girl will have written a book.

2.13.6 Imperfective-Continuous Aspect

Imperfective continuous aspect refers to the ongoing continuous situation during the reference time. Present imperfective continuous in Sindhi can be formed by aorist (old-present) base form followed by tensed-continuity marker “payO” which is further marked by gender and number in present and future tenses. However, in past tense only intransitive verbs are marked by gender and number while the auxiliary continuity marker remains un-marked. Future imperfective continuous is formed by pronominal suffixation of present-

participle (marked by number and gender) followed by future tense continuity marker (also marked by number and gender). Alternatively, like perfective construction imperfective-continuous is also formed by compound verb (conjunctive-participle+past-participle-Aux) in which past-participle-Aux form “rahaNra” is actually a continuity marker and is further marked by gender and number. Table III-49 shows different examples of imperfective-continuous formation in present, past and future tenses. In some dialects, instead of “payO” either “vETHO” or “payO-thO” markers are also used.

Table III - 49: Imperfective Continuous Constructions with “payO” Marker

Tense	Verb Type	Base Form	Example
Present	Transitive	Old-present-base + present tense-continuity marker	likHAN payO/ pal
	Intransitive		dORAN payO / pal
Past	Transitive	perfective-past-participle + past tense continuity marker	likHyO paE
	Intransitive		dORiyasi paE
Future	Transitive	Present-participle+pronominal-suffix + future tense continuity marker	likH-aNd-asi / -yasi payO/pal
	Intransitive		dOR-aNd-asi /-yasi payO/pal

Examples of imperfective continuous constructions with payO marker are shown below in (13a – 13d).

AUN kitAbu likHAN payO ----- (13a)
 I.Nom book.M.Sg.Nom write.OldPres.1P.Sg keep.Cont..Pres.M.Sg
 I am writing (a) book.

CHOkirl kitAbu likHE pal ----- (13b)
 girl.Nom book.M.Sg.Nom write.OldPres.3P.Sg keep.Cont..Pres.M.Sg
 Girl is writing (a) book.

CHOkirl kitAbu likHiyO paE ----- (13c)
 girl.Nom book.M.Sg.Nom write.Perf.Sg keep.Cont.Past
 Gir was writing (a) book.

AUN kitAbu likH-aNd-si payO ----- (13d)
 I.Nom book.M.Sg.Nom write-1P.M.Sg keep.Cont.Fut
 I will be writing a book.

As discussed above there is an alternate formation of imperfective continuous aspect. Table III-50 shows the examples of alternate formation in present, past and future tenses. Examples (14a – 14e) show that in case of present tense auxiliary holds number agreement only; however, in past and future tenses auxiliary holds number and gender agreement with main verb.

Table III - 50: Imperfective Continuous Constructions with Conjunctive Participle and “rah-iyO” marker

Tense	Verb Type	Base Form	Example
Present	Transitive	Conjunctive-participle + past-participle-aux + Old-present Aux	likHI rah-iyO/rah-I AhiyAN
	Intransitive		dORI rah-iyO/rah-I AhiyAN
Past	Transitive	Conjunctive-participle + past-participle-aux + past-Aux + pronominal suffix	likHI rah-iyO/rah-I hu-asi / hu-yasi
	Intransitive		dORI rah-iyO/rah-I hu-asi / hu-yasi
Future	Transitive	Conjunctive-participle + past-participle-aux + future-Aux + pronominal suffix	likHI rah-iyO/rah-I hUND-asi / hUND-iyasi
	Intransitive		dORI rah-iyO/rah-I hUND-asi / hUND-iyasi

CHOkirO kitAbu likHI-rahiyO AhE ----- (14a)
 boy.Nom book.M.Sg.Nom write.ConjPart- is.Aux.Pres.Sg

Boy is writing (a) book.

CHOkirl kitAbu likHI-rahI AhE ----- (14b)
 girl.Nom book.M.Sg.Nom write.ConjPart- is.Aux.Pres.Sg

Girl is writing (a) book.

CHOkirIUN kitAbu likHI-rahiyUN Ahen ----- (14c)
 boy.Nom book.M.Sg.Nom write.ConjPart- are.Aux.Pres.PI

Girls are writing (a) book.

CHOkirl kitAbu likHI-rahI hul ----- (14d)
 girl.Nom book.M.Sg.Nom write.ConjPart- was.Aux.Past.F.Sg

Girl was writing (a) book.

CHOkirA kitAbu likHI-rahiyA hUNDa ----- (14e)

boys.Nom book.M.Sg.Nom write.ConjPart-
Cont.M.PI will-be.Aux.Fut.M.PI
Boys will be writng (a) book.

2.13.7 Imperfective-Habitual Aspect

Imperfective habitual aspect expresses the occurrence of an event habitually. In Sindhi, present imperfective habitual is formed by present participle (marked by number and gender) form of verb followed by old present form of auxiliary “AhE” (marked by number). Past imperfective habitual has almost same structure as present imperfective habitual; the only difference is that instead of using old present form of auxiliary “AhE” the past participle form of auxiliary verb “AhE” with pronominal suffixation (hu-al / hu-yasi) is used. Future imperfective habitual also has same structure with pronominal suffixation of future auxiliary. Table III-51 shows different constructions of imperfective habitual aspect.

Examples (15a – 15c) show different sentences with imperfective habitual constructions given in Table IV-42 along-with agreement patterns.

Table III - 51: Imperfective Habitual Construction

Tense	Verb Type	Base Form	Example
Present	Transitive	Present-participle + Old-present Aux	likHa-NdO/dl AhyAN
	Intransitive		dORa-NdO/dl AhyAN
Past	Transitive	Present-participle + past-participle-aux + past-Aux + pronominal suffix	likHa-NdO/dl huasi/huyasi
	Intransitive		dORa-NdO/dl huasi/huyasi
Future	Transitive	Present-participle + past-participle-aux + future-Aux + pronominal suffix	likHa-NdO/dl hUND-asi/hUND-yasi
	Intransitive		dORa-NdO/dl hUND-asi/hUND-yasi

AUN	kitAbu	liikHaNdO	AhyAN ----- (15a)
I.1P.Sg.Nom	book.M.Sg.Nom	write.PresPart.M.Sg	is.Aux.OldPres.1P.Sg
I used to write (a) book. (Subject is Masculine)			
AUN	kitAbu	liikHaNdI	huyasi ----- (15b)
I.1P.Sg.Nom	book.M.Sg.Nom	write.PresPart.F.Sg	was.Aux.Past.1P.F.Sg
I used to write (a) book. (Subject is Feminine)			
AUN	kitAbu	liikHaNdO	hUNDasi ----- (15c)
I.1P.Sg.Nom	book.M.Sg.Nom	write.PresPart.M.Sg	will.Aux.Fut.1P.M.Sg
I will(habitual) write(habitual) (a) book. (Subject is Masculine)			

2.13.8 Combining Aspects

It is interesting to note that aspects can be combined and the construction will have properties of both combined aspects. For example, imperfective continuous and imperfective habitual can be combined and combination pattern includes continuity marker (payO) followed by habitual form of verb followed by tense auxiliary. “payO” is used here only as continuity marker and does not play any role in tense marking. Number and gender inflections take place for all components of combined aspect here (only exception is old-present auxiliary in present tense which is not marked by gender). Table III-52 shows different patterns of combined continuous-habitual aspect.

Alternate habitual continuous formation is also possible where habitual form of verb is followed by habitual form of continuity marker (rahu / rahaNra). Both main verb and continuity marker are marked by number and gender. Different examples of such constructions are shown in Table III-53.

Table III - 52: Continuous-Habitual Construction

Tense	Verb Type	Base Form	Example Sg.Masc / Sg.Fem
Present	Transitive	Continuity-Marker+Imperfective-Habitual + Old-present Aux	payO / pal likH-aNdO / -aNdl AhiyAN
	Intransitive		payO / pal dOR-aNdO / -aNdl AhiyAN
Past	Transitive	Continuity-Marker+Imperfective-Habitual + past-participle-aux + past-Aux + pronominal	payO / pal likH-aNdO / -aNdl hu-asi / -yasi
	Intransitive		payO / pal dOR-aNdO / -aNdl hu-asi / -yasi
Future	Transitive	Present-participle + past-participle-aux + future-Aux + pronominal suffix	payO / pal likH-aNdO / -aNdl huNd-asi / -iyasi
	Intransitive		payO / pal dOR-aNdO / -aNdl huNd-asi / -iyasi

Example sentences of continuous habitual patterns of Table III-52 are given in (16a – 16c). Gender and number agreement with feminine and masculine subjects can be seen in these examples.

CHOkirA kitAbu PayA likHaNdA ----- (16a)
 boys.M.Pl.Nom book.M.Sg.Nom keep.Cont.M.Pl write.PresPart.M.Pl

Ahen
 are.Aux.Pres.Pl
 Boys keep writing (a) book.

CHOkirUN kitAbu payUN likHaNdUN ----- (16b)
 girls.F.Pl.Nom book.M.Sg.Nom keep.Cont.M.Pl write.PresPart.M.Pl

huy-UN
 were.Aux.Past.3P.F.Pl
 Girls kept writing (a) book.

CHOkirUN kitAbu payUN likHaNdUN ----- (16c)
 girls.F.Pl.Nom book.M.Sg.Nom keep.Cont.M.Pl write.PresPart.M.Pl

hUND-UN
 were.Aux.Fut.3P.F.Pl
 Girls will keep writing (a) book.

Table III - 53: Alternate Continuous-Habitual Construction

Tense	Verb Type	Base Form	Example
Present	Transitive	Present-participle+ppCM + Old-present Aux	likH-aNdO / -aNdl rahaN-dO/-dl AhyAN
	Intransitive		dOR-aNdO / -aNdl rahaN-dO/-dl AhyAN
Past	Transitive	Present-participle + ppCM + past-Aux + pronominal suffix	likH-aNdO / -aNdl rahaN-dO/-dl huyasi
	Intransitive		dOR-aNdO / -aNdl rahaN-dO/-dl huyasi
Future	Transitive	Present-participle + ppCM + future-Aux + pronominal suffix	likH-aNdO / -aNdl rahaN-dO/-dl hUND-asi/-yasi
	Intransitive		dOR-aNdO / -aNdl rahaN-dO/-dl hUND-asi/-yasi

Examples (17a – 17c) show different patterns of alternate continuous habitual constructions.

CHOkirA kitAbu likHaNdA ----- (17a)
 boys.M.PI.Nom book.M.Sg.Nom write.PresPart.M.PI

rahaNdA Ahen are.Aux.Pres.PI
 stay.Cont.M.Sg

Boys use to keep writing (a) book.

CHOkiryUN kitAbu likHaNdYUN ----- (17b)
 girls.F.PI.Nom book.M.Sg.Nom write.PresPart.F.PI

rahaNdYUN hu-yUN were.Aux.Past.3P.F.PI
 stay.Cont.F.Sg

Girls used to keep writing (a) book.

CHOkiryUN kitAbu likHaNdYUN ----- (17c)
 girls.F.PI.Nom book.M.Sg.Nom write.PresPart.F.PI

rahaNdYUN hUNDyUN will-be.Aux.Fut.3P.F.PI
 stay.Cont.F.Sg

Girls will keep writing (a) book.

Continuous habitual pattern(s) shown in Table III-53 can also be preceded by “payO” marker; in this case “payO” is used as stress marker for continuity of the event (with number and gender inflections). Another possible present continuous habitual pattern can be seen in “likH-aNdO rahAN tHO” where pattern is “PresentParticiple+AoristContMarker+PresAux”.

2.14 Mood

Mood is grammatical reflection of the speaker's purpose in speaking (Kroeger, 2005). Bybee (1985:22) defines mood as an indication of what the speaker wants to do with the proposition. Palmer (1986:16) defines mood as a grammatical feature of verbs which is used to signal the modality; that is attitude of speaker towards what he is saying. Mood can be seen as a morphological variation which expresses or indicates the fact, desire, command or question etc. The verb mood can either be expressed by verbal inflections or using auxiliaries. J.S. Cole (2001) mentions four different moods in Sindhi namely subjunctive, presumptive, imperative, and counter-factual. The other mood categories also exist in Sindhi which include declarative or indicative, permissive, prohibitive, capacitive, suggestive, and compulsive moods. Following sections discuss these moods in detail.

2.14.1 Subjunctive Mood

Subjunctive mood generally expresses desires, hopes, wishes, conditions, feelings, opinions and imaginary events. Subjunctive is marked by old present form. Verb form must agree with subject in number and person.

Table III-54 shows different constructions of subjunctive mood. Sentences (18a – 18d) show different examples of subjunctive mood formation.

hU	dORE		----- (18a)
he/she	run.OldPres.Sg.3P		
If he/she runs / he/she shall run			
hU	dORaNdO	hujE	----- (18b)
he/she	run.OldPres.Sg.3P	might.Aux.OldPres.Sg.3P	
Perhaps he/she runs / he might be running			
Huni	dORiyO	hujE	----- (18c)
he/she.Obl.Sg	run.PastPart.Sg	might.Aux.OldPres.Sg.3P	
He/she may have run.			
Hu	dORE	hA	----- (18d)
he/she	run.OldPres.Sg.3P	might.Aux.Past	
He/she might had run / had he/she run.			

Table III - 54: Subjunctive Mood Constructions

Tense / Mood	Formation Base	Example
Old Present, Contingent Future Subjunctive	Old-Present form	hU likH-E (if he writes, he shall write)
Old-Present Habitual, Future Habitual / Subjunctive imperfective	Present participle + Old-Present form of Passive Aux huji	hU likH-aNdO hujE (perhaps he writes) (he might be writing)
Old-present- Perfective, Subjunctive Perfective	Perfective-past- participle + Old- Present form of Passive Aux huji	Hun-i likH-iyO huj-E (he may have written)
Past (conditional) / Subjunctive- Conditional, Contra factual	Old-present + Past- tense Aux (ha:)	hU likH-E hA (he might had written) (had he written)

2.14.2 Presumptive Mood

The form of verb that indicates presumption or supposition and is likely to be true. In Sindhi, presumptive mood is marked by present/past participles followed by present-participle form of auxiliary "huaNra" (to be). Consider the following example of imperfective-presumptive mood (19).

Hu	likH-aNd-O	hUNd-O	-----	(19)
he.Nom	write-PresPart-M.Sg	be.Aux.PresPart-M.Sg		
He would be writing.				

The verb form "likH-aNd-O hUNd-O" can also be said in future habitual-imperfective form. In the same way consider the perfective-presumptive mood of following example (20).

Huni	likH-iyO	hUNd-O	-----	(20)
he.Nom	write-PastPerf-M.Sg	be.Aux.PresPart-M.Sg		
He would have written.				

The verb forms "likH-iyO hUNd-O" in this case can also be said in future-perfective tense discussed above. "hUNdO" formation is ambiguous as it is not distinguishable with future auxiliary "hUNdO". Table III-55 shows different formations of presumptive mood. The "hUNdO" formation here is ambiguous as it is not distinguishable with future auxiliary "hUNdO".

Table III - 55: Presumptive Mood Formations

Tense / Mood	Formation Base	Example
Future Imperfective-Habitual / Imperfective-presumptive	Present-participle + PresPart Aux huaNra	likH-aNd-O hUNd-O
Future-Perfective / Perfective-presumptive	Past-Perfective-participle + PresPart Aux huaNra	likH-iyO hUNd-O

2.14.3 Imperative Mood

Imperative is used to express a command or a request. While comparing with moods discussed above imperative mood in Sindhi can be said an independent mood as it is not mixed up with tense-aspect system. Like English and Urdu, (Jafar Rizvi, 2007) Sindhi also uses verb stem as imperative form which may be followed by different intensifiers to express politeness (See section 1 for more discussion). Table III-56 shows different imperative forms of verb paRh-a (read).

Table III - 56: Imperative Mood Forms

Imperative Type	Example Imperative Stem of Verb paRh-a (1stPSG)
Rude / Frank	paRh-a
Formal	paRh-O (You Read Please)
Prohibitive	na paRh-a (Don't read)
Requested-Polite	paRh-a na / ba / ta (Please read)
Polite	paRh-iji (please you have to read)
Desiderative	paRh-ijiAni (please be written by you)

2.14.4 Declarative or Indicative Mood

Declarative mood describes the state of being something as a factual statement, declaration or indication. Like Urdu (Rizvi, 2007) in Sindhi declarative mood is marked by copula verbs “AhE”, “tHiyE” with argument structure like main verbs. For example, in the sentence “ahmed bImAr AhE” (ahmed is sick) the copula AhE has two arguments the subject (i.e. ahmed) and information about the subject (PREDLINK in LFG theory) i.e. “bImAr”.

2.14.5 Permissive, Prohibitive, Capacitive, Suggestive and Compulsive Moods

Permissive mood describes that subject permits the object to perform some action. In permissive mood, infinitive form of verb is followed by any of the inflected form of verb “DE” (give). Various forms of “DE” (DinU, DinI, DinAIN, etc.) act as permissive auxiliaries when preceded by infinitive. Sentences (21a – 21b) show examples of permissive construction.

mUN	CHOkirE-Khe	dORaNra	DinU	-----	(21a)
I.Obl	boy.Obl-Dat	run.Inf	gave.Sg		
I let boy to run.					

mUN	huna-Khe	mAnl	kHAiNra	dinI	-----	(21b)
I.Obl	him.Obl-Dat	meal.F.Sg .Nom	eat.Inf	gave.F.Sg		
I let him/her eat/take (his/her) meal.						

In prohibitive mood infinitive is either followed by postposition “kHAN” / “tAN” and past-participle form “rOki-yO” (stopped) of verb rOki (stop) acting as prohibitive marker (instead of “rOki-yO”, “jHali-yO” is also used with same context) or just negating the permissive (i.e. infinitive preceded by negation auxiliary “na”). An example is shown in (22).

mUN	CHOkirE-kHE	dORaNra-kHAN	jHaliyO	-----	(22)
I.Obl	boy.Obl-Dat	run.Inf-from	stop.PastPart.Sg.M		
I stopped the boy to run /from running.					

Capacitive mood expresses the capability of performing the action. Verb form in this mood is a compound verb which is formed by conjunctive participle followed by different forms of modal verb “sagHaNra” (to be able to do). Sentence (23) given below is an example of capacitive mood construction.

CHOkirO	gADI	halAE-sagHE	thO -----	(23)
boy.Nom	car.Nom	drive.ConjPart-can	be.Pres.Sg.M	
Boy can drive the car.				

Suggestive mood expresses suggestion or advice to a subject. The verbs “kHapE” and “gHurjE” follow the infinitive verb to form suggestive mood. Sentence (24) shows an example of suggestive mood.

CHOkirE-kHE	gADI	halAiNra	gHurjE/kHapE-----	(24)
Boy.Obl-dat	Car.Nom.F.Sg	drive.Inf-from	need / require	
The boy should drive the car.				

Compulsive mood is formed by future participle followed by different forms of auxiliary verb “payO”. Example is given in sentence (25).

CHOkirE-kHE	gADI	halAiNrl	pal -----	(25)
Boy.Obl-dat	Car.Nom.F.Sg	drive.FutPart.F.Sg	keep.Aux.Past.F.Sg	
Boy had (compulsive) to drive the car.				

Table III-57 shows mood patterns, markers and examples of different moods discussed above.

Table III - 57: Permissive, Prohibitive, Capacitive, Suggestive, and Compulsive Mood Constructions

Mood	Pattern	Auxiliary Maker(s)	Example
Permissive	Infinitive + DE-Inflections	D-in-U, D-in-A, D-in-I, D-ini-yUN, D-in-A-IN, D-in-A-UN	likH-aNra dinU (permitted to write)
Prohibitive	Infinitive + kHAN/tAN + rOki / jHali Inflections	rOki-yO, rOki-yA, rOki-yUN, rOk-I	likH-aNra kHAN rOki-yO (did not allow to write)
	na + Permissive Construction	Na	na likH-aNra dinU (did not permit to write)
Capicitive	Conjunctive-Participle + SagHa Inflections	sagH-E, sagH-AN, sagH-ani, sagH-UN	likH-I sagH-E (can write)
Suggestive	Infinitive + gHuri / kHapa Inflections	gHuri-ij-E, gHuri-ij-ani, kHap-E, kHap-ani	likH-aNra gHuri-ij-E (must write)
Compulsive	Future-participle + payO Inflections	pa-yO, pa-yA, pa-l, pa-yUN	likH-aNr-U pay-O (had to write)

2.15 Pronominal Suffixes

Sindhi is one of few Indo-Aryan languages (along-with Kashmiri, Punjabi and Seraiki) with pronominal suffixes. Pronominal suffix is a morphological change in a word which reflects the pronoun along with existing word class. For example, the noun “puTa” (son) can have pronominal suffix “-mi” and the word “puta-mi” (my son) has pronoun “my” and noun “son”. It may also be noted that syntactic function of word remains same i.e. after inflection word remains in same syntactic / part of speech class. Pronominal suffixes are usually declined for person and number and rarely for gender and are added to nouns, verbs and postpositions. Table III-58 shows different pronominal suffixes used with nouns.

Table III- 58: Pronominal Suffixes Used with Nouns

Word	Sg	Pl	Example
puTu (son)	-mi	-UN	puTu-mi (my son), puTu-UN (our son)
	-EN	-va	puTu-EN (your.sg son) puTu-va (your.pl son)
	-si	-ni	puTu-si (his son) puTu-ni (their son)
puTa (sons)	-mi	-UN	puTa-mi (my sons), puta-UN (our sons)
	-EN	-va	puTa-EN (your.sg sons) puTa-va (your.pl sons)
	-si	-ni	puTa-si (his sons) puTa-ni (their sons)

Examples of pronominal suffixes given in Table III-58 are shown in sentences (26a – 26g).

puTu-mi acHE tHO ----- (26a)
 son.Sg-1P.Sg come.OldPres.Sg be.Aux.Pres.M.Sg
 My son comes.

puTu-UN son.Sg-1P.PI Our son comes.	acHE come.OldPres.Sg	tHO be.Aux.Pres.M.Sg	----- (26b)
puTu-EN son.Sg-2P.Sg Your son comes.	acHE come.OldPres.Sg	tHO be.Aux.Pres.M.Sg	----- (26c)
puTu-si son.Sg-3P.Sg His son comes.	acHE come.OldPres.Sg	tHO be.Aux.Pres.M.Sg	----- (26d)
puTa-UN son.PI-1P.PI Our sons come.	acHani come.OldPres.PI	tHA be.Aux.Pres.M.PI	----- (26e)
puTa-va son.PI-2P.PI Your.PI sons come.	acHani come.OldPres.PI	tHA be.Aux.Pres.M.PI	----- (26f)
puTa-si son.PI-3P.Sg His sons come.	acHani come.OldPres.PI	tHA be.Aux.Pres.M.PI	----- (26g)

Pronominal suffixes for auxiliary verbs may vary depending on the tense. For example, present tense auxiliary verbs “AhE” (is) will have morphological changes caused by pronominal suffixes shown in Table III-59. It may also be noted that there are irregular inflections like “AhE” to “athHal”. The past tense auxiliary verb “huO” (was) will have different inflections and additionally will be marked by gender. Table III-60 shows different pronominal suffixes (with inflections) of past tense auxiliary “huO”. Example sentences with pronominal suffixes with present tense auxiliary AhE are given in (27a – 27c) whereas examples of pronominal suffixes with past tense auxiliary “huo” are given in (28a – 28c).

Table III - 59: Pronominal Suffixes used with Present Tense Auxiliary “AhE”

Person	Sg	Pl	Sg	Pl	Example	
1st	AhE	Ahin	-mi	-UN	AhE-mi, atha-mi (I have sg)	atha-UN (We have pl)
2nd			-I	-va	AhE-I, atha-I (You have)	AhE-va, atha-va (You have pl)
3rd			-si	-ni	AhE-si, atha-si (he has)	AhE-ni, atha-ni (they have)

KitAbu	AhE-mi	----- (27a)
Book.Sg	is.Aux.Pres-1P.Sg	
I have (a) book.		

kitAbu	AhE-va	----- (27b)
Book.Sg	is.Aux.Pres-2P.PI	
You.PI have (a) book.		

kitAbu AhE-ni ----- (27c)
Book.Sg is.Aux.Pres-3P.PI
They have (a) book.

Table III- 60: Pronominal Suffixes of Past Tense Auxiliary "hu-O"

Person	Masculine				Feminine			
	Singular		Plural		Singular		Plural	
	Affix	Example	Affix	Example	Affix	Example	Affix	Example
1 st	-mi	hu-O-mi	-A-sIN	hu-A-sIN	- yami	hu-I-yami	-UN-sIN	hu-UN-sIN
2 nd	-I	hu-O-I	-ava	hu-ava	-yal	hu-yal	- yava	hu-yava
3 rd	-si	hu-O-si	-ani	hu-ani	-yasi	hu-yasi	-yani	hu-yani

kitAbu	huO-mi	----- (28a)
Book.Sg	was.Aux.Past-1P.Sg	
I had (a) book.		

kitAbu	hu-yava	----- (28b)
Book.Sg	was.Aux.Past-2P.F.Pl	
You.Pl.F had (a) book.		

kitAbu hu-yani ----- (28c)
Book.Sg is.Aux.Pres-3P.F.Pl
They.F had (a) book.

Pronominal suffixation with auxiliary verbs also takes place with main verbs. For example, the intransitive verb “acHa” (come) will have different pronominal suffixes for different tenses. For unspecified or old present (see discussion on tense and aspect in section 2.13) examples are given in Table III-61. For past tense form “AyO” of verb “acHa” pronominal suffix form will be “A-yus-i” (I came). Examples are shown in sentences (29a – 29b).

Table III- 61: Pronominal Suffixes in Old-Present/Aorist Verb Forms

Person	Masculine / Feminine			
	Singular		Plural	
	Affix	Example	Affix	Example
2 nd	-i	achAN-i	-va	achAN-va
3 rd	-si	achAN-si	-ni	achAN-ni

acHAN-va	tHO	----- (29a)
come.OldPres.1P.Sg-2P.PI	be.Aux.Pres.M.Sg	
I(Masc) come to you.PI.		

AyO-va come.PastPart-2P.PI I came to you.PI	huO-mi was.Aux.Past-1P.Sg	----- (29b)
---	------------------------------	-------------

Pronominal suffixation in postpositions also takes place by using “-mi”, “-i”, and “-si” in singular and “-UN”, “-va”, and “-ni”, in plural forms. Examples of postpositional pronominal suffixation are shown in Table III-62. Some postpositions may also have number marking like postposition “jehRO” will also be marked by number while having pronominal suffixation (number marking is not shown in table). Sentences (30a – 30d) show different examples of pronominal suffix patterns given in Table III-62.

Huni he.Obl	khE-mi Dat-1P.Sg.Obl	kitAbu book.Sg.M	dinU give.PastPart.M.Sg	-----	(30a)
mUN I.Obl	khE-si Dat-3P.Sg.Obl	kitAbu book.Sg.M	dinU give.PastPart.M.Sg	-----	(30b)
CHOkirO boy.Nom.M.Sg	sANra-mi with-1P.Sg.Obl	AyO Came.PastPart.Sg.M	-----	-----	(30c)
hU he.Obl	jehRO-mi like-1P.Sg.Obl	AhE is.Aux.Pres.Sg.M	-----	-----	(30d)
	He is like me.				

Table III - 62: Pronominal Suffixation on Postpositions

Postposition	Person				
		Singular		Plural	
		Affix	Example	Affix	Example
kHE	1 st	-mi	kHE-mi	-	-
	2 nd	-	-	-	-
	3 rd	-si	kHE-si	-ni	kHE-ni
sANra	1 st	-mi	sANra-mi	-UN	sANra-UN
	2 nd	-EN	sANra-EN	-va	sANra-va
	3 rd	-si	sANra-si	-ni	sANra—ni
jehRO	1 st	-mi	jehRO-mi	-UN	jehRO-UN
	2 nd	-EN	jehRO-EN	-va	jehRO-va
	3 rd	-si	jehRO-si	-ni	jehRO-ni

2.15.1 Double Pronominal Suffixation

Double pronominal suffixes are also common in Sindhi. In double suffixation, inflectional change reflects two pronouns mostly first and second or first and third person pronouns. Double pronominal suffixation is only used with verbs. Examples of double pronominal suffixation of past tense form of

verb “likHu” (write) are shown in Table III-63 and Table III-64. Example sentences showing double pronominal suffixation are given in (31a – 31c).

Table III- 63: Double Pronominal Suffixation Examples of Verb “liKhu” 1st Person – 2nd Person and 1st Person – 3rd Person

Affix-1 Number	Person	Singular Affix-2			Plural Affix-2		
		Affix-1	Affix-2	Example	Affix-1	Affix-2	Example
Singular	2nd	-mAN	-i	likHiyO-mAN-i I wrote you	-mAN	-va	likHiyO-mAN-va I wrote you
	3rd	-mAN	-si	likHiyO-mAN-si I wrote him	-mAN	-ni	likHiyO-mAN-ni I wrote them
	2nd	-sUN	-i	likHiyO-sUN-i We wrote you (Sg)	-sUN	-va	likHiyO-sUN-va We wrote you (Pl)
Plural		-sUN	-si	likHiyO-sUN-si We wrote him	-sUN	-ni	likHiyO-sUN-ni We wrote them

Table III- 64: Double Pronominal Suffixation Examples of Verb “liKhu” 3rd Person – 1st Person, 3rd Person – 2nd Person and 3rd Person – 3rd Person

Person	Singular			Plural		
	Affix-1	Affix-2	Example	Affix-1	Affix-2	Example
1st	-A-IN/OIN	-mi	likHiy-AIN-mi He wrote me	-	-	-
2nd	-A-IN / OIN	-i	likHiy-AIN-i He wrote you	-A-IN	-va	likHiy-AIN-va He wrote you
3rd	-A-IN / OIN	-si	likHiy-AIN-si He wrote him	-A-IN	-ni	likHiy-AIN-ni He wrote them

KHatu likHiyO-mAN-va ----- (31a)
 Letter.Nom.Sg.M write.PastPart.Sg-1P.Sg-2P.PI
 I wrote you (a) letter.

KHatu likHiyO-sUN-si ----- (31b)
 Letter.Nom.Sg.M write.PastPart.Sg-1P.PI-3P.Sg
 We wrote him (a) letter.

KHatu likHiyO-IN-ni ----- (31c)
 Letter.Nom.Sg.M write.PastPart.Sg-3P.Sg-3P.PI
 He wrote them (a) letter.

2.16 Postpositional Suffixes

Sindhi also has postpositional suffixes usually in nouns, and adverbs.

Examples of postpositional suffixes are shown in Table III-65 and example sentences are shown in (32a – 32b).

tUN gHar-AN acHu ----- (32a)
 you.Sg home-from.Abl come.Sg.Imp
 You come from home.

tUN BAra-EN acHu ----- (32b)
 you.Sg Child.PI-with.Part come.Sg.Imp
 You come with children.

Table III - 65: Postpositional Suffixes

POS	Suffix	Example	Equivalent
hETHa (Adverb) Down / Below	-AN	hETH-AN	hETHa kHAN
		from down / below	
gHaru (Noun) Home	-AN	gHar-AN	gHaru mAN / kHAN
		from home	
BAru (Noun) Child	-EN	BAr-EN	BAr-ani sAN
		with children	

3. SYNTAX

Syntax deals with phrase and sentence structures and includes nominal and verbal elements. Along-with nouns nominal elements may also include pronouns, adjective phrases, postpositional phrases, participle phrases, and relative clauses. Noun phrases can either be marked by case at morphology level or at syntax level. Verbal arguments are defined by verb subcategorization frames. Details of nominal and verbal syntax of Sindhi are discussed in subsequent sections.

3.1 Noun Phrase in Sindhi

Phrase is a group of related words which forms a component of a clause; a noun phrase is a phrase which plays the role of noun (subject, object, indirect object, and predlink etc.) in a sentence. Noun phrase in Sindhi may have simple or complex constructions. We will start our discussion with simple constructions. Consider following sentences (Note that Aorist used in gloss is equivalent of OldPres).

CHOkirO	likHE	thO	-----	(33a)
boy.Nom	write.Aorist	be.Sg.M.Pres.Aux		
The boy writes.				
CHOkirE	likHiyO	huO	-----	(33b)
boy.Obl	write.PastPart	be.Sg.M.Past.Aux		
The boy wrote.				
ali	likHE	thO	-----	(33c)
Ali.Nom	write.Aorist	be.Sg.M.Pres.Aux		
Ali writes.				
Ali	likHiyO	huO	-----	(33d)
Ali.Nom	write.PastPart	be.Sg.M.Past.Aux		
Ali wrote.				

Various constituents in these sentences include noun, verb and auxiliary verb; the phrase structure rule for above sentences will be.

$S \rightarrow NP\ V\ (VAUX);$
 $NP \rightarrow N.$

Where S represents sentence and the rule says that a sentence can be defined as noun phrase (NP) followed by a verb (V) and an optional auxiliary verb (VAUX). “CHOkirO” (nominative common noun form), “CHOkirE” (oblique common noun form) and “Ali” (proper noun) are noun phrase examples in above sentences. As nouns are marked by number gender and case so is the case with noun phrase. NP construction may also include standalone pronouns. Consider following examples.

hU he.Nom.3P.M.Sg He writes.	likHE write.Aorist	thO be.Pres.Aux	----- (34a)
Hina he.Obl.3P.M.Sg He wrote.	likHiyO write.PastPart		----- (34b)
ihO it.Nom.3P.M.Sg It writes.	likHE write.Aorist	thO be.Pres.Aux.Sg.M	----- (34c)
inhE it.Obl.3P.Sg It wrote.	likHiyO write.PastPart		----- (34d)
jEkO which/who.Nom.rel.Sg.M Who writes.	likHE write.aorist	thO be.Pres.Aux.Sg.M	----- (34e)
kEra who.whp Who writes.	likHE write.aorist	thO be.Pres.Aux.Sg.M	----- (34f)
CHA what.whp What writes?	likHE write.aorist	thO be.Pres.Aux.Sg.M	----- (34g)

kehRO Who.indef.Nom.M.Sg Which/Who writes?	likHE write.aorist	thO be.Pres.Aux.Sg.M	----- (34h)
pANra he.refl He (himself) writes.	likHE write.aorist	thO be.Pres.Aux.Sg.M	----- (34i)

Examples (34e) and (34f) are sentences with personal pronouns, (34g) and (34h) are with demonstrative pronoun NPs. Rest of the sentences are with relative, wh-pronoun, indefinite pronoun and reflexive pronoun NPs. To cover the above examples, NP can be redefined as given below.

NP → N | PProng | NPPron;
 NPPron → DPron | WhPron | IndPron | RefPron.

Above rule defines NP as noun (N) or personal pronoun (PProng) or NPPron which is further defined as either of four different pronoun types shown in above examples. Personal pronoun is not included in NPPron group due to its different syntactic properties which don't allow personal pronouns to be combined with other NP components like nouns and adjectives discussed in subsequent sections.

3.1.1 Complex NP Constructions

NP definition discussed above only includes either single noun or pronoun. However, NP can have combination of different NP components which include pronoun-noun, adjective-noun, and pronoun-adjective-noun constituents. These combinations are further complicated by coordination, postpositional phrases and relative clauses. Various such constructions are discussed below.

Consider following examples:

ihO it.demon.Nom.M.Sg That boy writes.	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	---- (35a)
inhE it.demon.Obl.Sg That boy wrote.	CHOkirE boy.obl	likHiyO write.PastPart		---- (35b)
jEkO which/who.rel The boy who writes.	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	---- (35c)
kEra Who/which.whp Which boy writes?	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	---- (35d)
CHA what.whp Does the boy write?	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	---- (35e)
kO someone.ind.Nom.Sg.M Some boy writes.	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	---- (35f)
pANra himself.refl The boy writes himself.	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	---- (35g)

Four different pronoun types which are defined as NPPron above are combined with common noun "CHOkirO". NP therefore can be redefined with little bit alteration where NPPron is optionally followed by a noun (usually common noun).

NP → N | PPrön | NPPron (N);
 NPPron → DPrön | WhPron | IndPron | RefPron.

In adjective noun combination adjectives, which modify nouns are placed before nouns. Consider following sentences.

sutHO good.Nom A good boy writes.	CHOkirO boy.nom	likHE write.aorist	thO be.Pres.Aux.Sg.M	----- (36a)
---	--------------------	-----------------------	-------------------------	-------------

suhiNrE CHOkirE likHiyO ----- (36b)
 beautiful.Obl boy.obl write.pastpart
 A beautiful boy wrote.

kArA bHOliRA dORani thA ----- (36c)
 black.Nom.Pl monkeys.Nom run.Aorist.Pl be.Pres.Aux.PI.M
 Black monkeys run.

cCHAR Bara dORani thA ----- (36d)
 four.Nom children.Nom run.Aorist.Pl be.Pres.Aux.PI.M
 Four children run.

NP in above four sentences is combination of adjectives followed by noun. NP definition may now contain this combination possibility too as given below.

$NP \rightarrow (ADJ) N \mid P\text{Pron} \mid NP\text{Pron} (N);$

It may also be noted that standalone adjective can function as NP and it can appear without noun; NP definition of noun-adjective combination with this optional adjective-only construction is now modified as shown below.

$NP \rightarrow (ADJ) N \mid ADJ \mid P\text{Pron} \mid NP\text{Pron} (N);$

These adjective combinations can be coordinated with multiple adjectives which form adjective phrases. Theoretically any number of adjective phrases can precede noun. Consider following sentences which include multiple adjectives before nouns. This adjective coordination either include two adjectives separated by a coordinative conjunction (see 37a and 37b) or a list of adjectives delimited by space or comma followed by coordinated conjunction and single adjective (see 37c).

sutHO aiN hOshiyAr CHOkirO ----- (37a)
 good.Nom.M.Sg And intelligent.Nom boy.Nom
 likHE thO

write.aorist be.Pres.Aux.Sg.M
 A good and intelligent boy writes.

suhiNrE aiN sutHE CHOkirE likHiyO ----- (37b)
 beautiful.Obl.M.Sg and good.Obl.M.Sg boy.Obl write.pastpart
 A beautiful and good boy wrote.

naNdHA kArA aiN suhiNrA bHOliRA ----- (37c)
 small.Nom.M.Pl black.Nom.M.Pl and beautiful.Nom.M.Pl Monkeys.Nom.M

dORani thA
 run.Aorist.Pl are be.Pres.Aux.PI.M
 Small, black and beautiful monkeys run.

NP definition is now redefined with ADJP replacing ADJ. ADJP is defined as a single adjective optionally followed by zero or any number of adjectives ADJ* followed by conjunction and adjective.

$NP \rightarrow ADJP^* N \mid ADJP \mid P\text{Pron} \mid NP\text{Pron} (N);$
 $ADJP \rightarrow ADJ (ADJ^* CONJ ADJ).$

3.1.2 Postpositional Phrases and Relative Clauses

Postpositional phrase may also modify NPs. Consider following examples in which postpositional phrase is used in NP.

mEZa tE vEtHala CHOkirO ----- (38)
 table.Obl on sit.pastpart boy.Nom
 A boy sitting on the table.

Here “mEZa tE” is a postpositional phrase followed by past participle “vEtHala” (used as an adjective here). Postpositional phrase can be defined as a noun followed by a postposition.

$PPP \rightarrow N PP.$

In NP definition, postpositional phrase can optionally occur before ADJP as postpositions never directly modify noun phrases. Postpositional phrase modifies the adjective phrase which then modifies the noun phrase. It may also be noted that in this case adjective phrase is past participle clause used as participle relative clause which in turn modifies the noun phrase. There can be any number of postpositional phrases so the NP definition is modified.

$NP \rightarrow PPP^* ADJP^* N | ADJP | PPr\acute{o}n | NPPr\acute{o}n (NC);$
 $ADJP \rightarrow ADJ (ADJ^* CONJ ADJ).$
 $PPP \rightarrow N PP.$

Along-with all constituents discussed above noun phrase may also contain relative clause. Relative clauses are clauses which start with relative pronouns (“jEkO”, “jehRO” etc.) used to define the NP further. For example, in following sentence the clause “jEkO kArO AhE” is a relative clause which provides more information about the head noun of NP.

mEZa	tE	vEtHala	BHOliRO	jEkO -----	(39)
table.Obl.F.Sg	On	sit.PastPart	monkey.Nom.M.Sg	which.rel.M.Sg	
kArO	AhE				
black.Nom.M.Sg	be.Cop.Sg				

Monkey sitting on the table which is black.

NP with optional relative clause is defined below. CPRel will define relative clause (See Chapter VI for further discussion).

$NP \rightarrow \{ PPP^* (ADJP) N | ADJP | PPr\acute{o}n | NPPr\acute{o}n (NC) \} (CPRel).$
 $ADJP \rightarrow ADJ (ADJ^* CONJ ADJ).$
 $PPP \rightarrow N PP.$

3.2 Case Marking

Nouns are marked by case and so is the situation with noun phrase.

Morphological case of nouns is discussed in section 2.3. Nominative case is without any case marker and is known as default nominative morphological form of noun. However, further case marking takes place in oblique form of nouns which is base form for the syntactic case marking. Subsequent sections discuss different cases of nouns along with their case markers and syntactic constructions.

3.2.1 Nominative Case

Nouns (or noun phrases) are said to be in nominative case if they are in nominative morphological form and there is no case marker. Nominative nouns can appear in a sentence either as a subject or complement. Complements in LFG can either appear as PREDLINK or XCOMP (Dalrymple, Dyvik, & King, 2004).

Consider following examples.

CHOkirO likHE thO ----- (40a)

boy.Nom.M write.Aorist.Sg be.Pres.Aux.Sg.M

Boy writes.

bHOliRA dORani thA ----- (40b)

Monkeys.Nom.M run.Aorist.PI be.Pres.Aux.PI.M

Monkeys run.

Bara dORani thA ----- (40c)

children.Nom.M run.Aorist.PI be.Pres.Aux.PI.M

Children run.

CHOkirO BAru AhE ----- (40d)

boy.Nom.M child.Nom.Sg.M be.Cop.Sg

Boy is a child.

Nominative is default case for nouns in nominative form (without any inflection); however other inflections (number and gender) still take place. In above examples nouns “CHOkirO” (boy), “bHOliRA” (monkeys) and “BAru” (child) are in nominative case. In (40d) the noun “BAru” (child) is used as a complement of “CHOkirO” and can be treated as PREDLINK of subject “CHOkirO”.

3.2.2 Obliqueness and Ergativity

As discussed in section 2.3.3 during case formation when nouns are followed by postpositions or case markers they appear in oblique form. This oblique form is further marked by number and gender. Table V – 1 shows different oblique forms of common noun “CHOkirrO” (boy). Examples are shown in (41a – 41d).

Table III - 66: Oblique forms of common noun ‘CHOkirO’

Gender	Number	Nominative	Oblique
Masc	Singular	CHOkirO	CHOkirE
	Plural	CHOkirA	CHOkirani
Fem	Singular	CHOkirl	CHOkirla
	Plural	CHOkiryUN	CHOkirlani

CHOkirE tE ----- (41a)
 boy.M.Sg On
 On the boy.

CHOkirani tE ----- (41b)
 boy.M.Pl on
 On the boys.

CHOkirla tE ----- (41c)
 girl.F.Sg on
 On the girl.

CHOkirlani tE ----- (41d)
 girl.F.Pl on
 On the girls.

Ergativity refers to the case of nouns expressing the role of an actor or agent used as a subject in the sentence. In Urdu/Hindi and Punjabi “nE” is used as an ergative case marker. However, in Sindhi we do not have a separate ergative case marker and ergativity is reflected by oblique form of nouns. Sentence (42a) shows example of ergative case marking in Hindi/Urdu and (42b) shows an equivalent Sindhi sentence where oblique form is used without any case marker.

laRkE-nE	sEb	kHAyA	-----	(42a)
boy.M.Sg-Erg	apple.nom	eat.Perf		
Boy ate an apple.				

CHOkirE	sUfu	kHAdHO	-----	(42b)
boy.M.Sg.Obl	apple.nom	eat.Perf		
Boy ate an apple.				

3.2.3 Accusative Case

Accusative case in Sindhi is marked by postposition “kHE” when noun or noun phrase is direct object in oblique form. Consider following examples of accusative case.

CHOkirO	kutE=kHE	mArE	thO	-----	(43a)
boy.Nom	dog.Obl.M.Sg=Acc	kill.Aorist.Sg	be.Pres.Aux.Sg.M		
The boy kills/beats the dog.					

CHOkirE	kutE=kHE	mAryO	-----	(43b)
boy.Obl	dog.Obl.M.Sg=Acc	kill.PerfPart		
The boy killed/beat the dog.				

CHOkirO	kutE=kHE	MArINdO	-----	(43c)
boy.Nom	dog.Obl.M.Sg=Acc	kill.Fut		
The boy will kill/beat the dog.				

(43a) is an example of present tense sentence and the subject “CHOkirO” is in nominative case, “kutE” (oblique form of “kutO”) is followed by case marker “kHE” and is in accusative case. In (43b) “CHOkirE” is oblique form of

“CHOkirO” and sentence is in past tense with perfective aspect which requires subject to be in oblique form; this reflects the ergativity of subject as discussed above.

3.2.4 Differential Object Marking

Direct object can either be marked by accusative case marker or remain un marked in nominative case depending on the type of object. This is called differential object marking. Differential object marking phenomenon in Urdu-Hindi is also common (Butt, (1993), (Masica, 1982), (Mohanan, 1995).

Consider following examples.

CHOkirO Bakirla=kHE
boy.Nom goat.Obl.F.Sg=Acc
The boy beats the goat.

CHOkirO mAnla=kHE*
boy.Nom meal.Obl.F.Sg=Acc*
The boy eats a bread/meal.

CHOkirO mAnl
boy.Nom meal.Nom.F.Sg
The boy eats a bread/meal.

mArE tHO ----- (44a)
beat.OldPres be.Pres.Aux.Sg.M

kHAE tHO ----- (44b)
eat.OldPres be.Pres.Aux.Sg.M

kHAE tHO ----- (44c)
eat.OldPres be.Pres.Aux.Sg.M

In (44a) direct object ‘Bakirl’ is marked by accusative case marker ‘kHE’. However, in (44b) use of ‘kHE’ marker is quite odd and this pattern not used is Sindhi. (44c) shows an example direct object ‘mAnl’ (meal/bread) without any case marker (it is in default nominative case). This phenomenon can also be observed in (42b) where object ‘sUfu’ (apple) is unmarked and subject ‘CHOkirO’ (boy) reflects ergativity in oblique form. This pattern is very common in Sindhi where oblique subjects appear with perfective forms of verb.

3.2.5 Dative Case

Noun phrase which acts as an indirect object (OBJ2) is marked by dative case. “kHE” is also used as dative case marker. Consider following sentences.

CHOkirO Ali=kHE boy.Nom Ali.M.Sg=Dat The boy gives a book to Ali.	kitAbu DE tHO ----- (45a) book.Nom.M.Sg give be.Pres.Aux.Sg.M
---	--

Ali ChokirE=khE Ali.M.Sg Boy.Obl=Dat Ali gave a book to the boy.	kitAbu DinU ----- (45b) book.Nom.M.Sg gave.PastPart.M.Sg
--	---

CHOkirE Ali=khE Boy.Obl Ali=Dat The boy gave a book to Ali.	kitAbu DinU ----- (45c) book.Nom.M.Sg gave.PastPart.M.Sg
---	---

In (45a) “CHOkirO” is a subject in nominative case, “kitAbu” is direct object in nominative case, and Ali is indirect object with dative case. “kHE” can either mark dative or accusative cases and the case is identified by verbal subcategorization frame (see Section 2). Sometimes subject is also marked by dative case. Following examples (46a – 46c) show subjects with dative case.

CHOkirE=kHE bukHAr boy.Obl=Dat fever.Nom The boy has fever.	AhE be.Cop.Sg ----- (46a)
---	---------------------------------

Ali=kHE iskUlu Ali=Dat school.Nom.M.Sg Ali has to go to school.	vanNhNrU AhE go.inf.Sg be.Cop.Sg ----- (46b)
---	--

mUN=kHE KHayAlu me.Obl=Dat idea.Nom.M.Sg I had an idea.	AyO came.M.Sg ----- (46c)
---	---------------------------------

3.2.6 Participant Case

Participant case is formed when subject is animate noun and the other NP is also animate noun. Postposition “sAN” is used to mark participant case; presence of subject is mandatory in sentence. Other animate noun is experiencer/participant of the action and must be in oblique form. In following sentences “Ali”, “mUN” and “huna” are examples of participant case.

mUN Ali=sAN GALhAyO ----- (47a)
me.Obl Ali.M.Sg=Part talk.PastPart
I talked to/with Ali.

CHOkirO	mUN=sAN	dORyO	----- (47b)
boy.Nom	me.Obl=Part	run.PastPart	
The boy ran with me.			

CHOkirE	huna=sAN	likHiyO	-----	(47c)
boy.Obl	he.Obl.Sg=Part	write.PastPart		
The boy wrote with him.				

The noun in participant case is treated as an ADJUNCT instead of OBJ (object) or OBJ2 (indirect object). Consider following sentences.

mUN Ali=sAN sUfu khAdHO ----- (48a)
me.Obl Ali=Part apple.Nom Ate
I ate an apple with Ali.

mUN	Ali=sAN	CHOkirE=khE	kitAbu	DinO	----- (48b)
me.Obl	Ali=Part	boy.Obl=Dat	book.Nom	gave	
I gave a book to a boy with Ali.					

In (48a) “mUN” is subject, “sUfu” is object, Ali is neither indirect object nor object just the participant of subject. In (48b) subject (mUN), direct object (kitAbu), indirect object (CHOkirE) are there and ‘Ali’ is just a participant of subject and therefore is ADJUNCT in both sentences.

3.2.7 Instrumental Case

Instrumental case is also marked by postposition “sAN”. Inanimate nouns in oblique form followed by postposition “sAN” are said to be in instrumental case. NP in instrumental case is also considered ADJUNCT in a sentence like participant case discussed above. In following sentences (49a – 49c), “ChamChE”, “kAtI”, and “saUNTI” are examples of instrumental case.

mAu	ChamChE=sAN	bAru=kHE	----- (49a)
mother.Obl	spoon.Obl.M.Sg=Inst	child.M.Sg=Dat	

khAdHO	khArAyO
food.Nom.M.Sg	eat.Caus1

Mother caused the child to eat with a spoon.

mUN	kAtla=sAN	sUfu	----- (49b)
me.Obl	knife.F.Sg=Inst	apple.M.Sg.Nom	

kaTyO
cut.PastPart.M.Sg

cut an apple with knife.

mUN	saUNTIa=sAN	gaDahu=kHE	----- (49c)
me.Obl	cane.Obl.F.Sg=Inst	donkey.M.Sg =acc	

mAriyO
beat.PastPart.M.Sg

I beat the donkey with cane.

3.2.8 Locative Case

Locative case indicates a location. Locative case is marked when noun phrase in oblique form is followed by a postposition which acts as a locative case marker. In Sindhi locative case marker postpositions are “mEN”, “tE” and “tAIN”. In examples (50a – 50b) “kamrE” and “gHaru” are examples of locative case.

CHOkirO kamrE=mEN AhE ----- (50a)
 boy.Nom room.Obl.M.Sg=Loc be.Cop.Sg
 The boy is in the room.

Ali kursla=tE likHE thO ----- (50b)
 Ali.M.Sg chair.Obl.F.Sg=Loc write.Aorist.Sg be.Pres.Aux.Sg.M
 Ali writes on the chair.

3.2.9 Ablative Case

Ablative case indicates motion away or movement away from something (NP). In Sindhi ablative case is either represented by morphological form or ablative case marker “kHAN”. Consider following examples.

mUN CHOkirE=kHan kitAbu vartO ----- (51a)
 me.Obl boy.Obl=Abl book.Nom.M.Sg took.PastPart.M.Sg
 I took a book from boy.

Ali gHaru=kHan bHaGO ----- (51b)
 Ali.Nom.M.Sg home.M.Sg=Abl run.PastPart.M.Sg
 Ali ran from home.

Ali gHarAN bHaGO ----- (51c)
 Ali.Nom.M.Sg home.Abl run.PastPart.M.Sg
 Ali ran from home.

(51a) and (51b) show the ablative case marking with “kHAN” marker where noun must be in oblique form. (51c) shows the ablative morphological form of “gHaru” no case marker is used here but morphological form. NP in ablative case is also considered ADJUNCT in a sentence.

3.2.10 Agentive Case

Agentive case in Sindhi is marked with postposition “kHAN”, i.e. “kHAN” not only acts as an ablative case marker but agentive case marker as well. Agentive case is formed during passivization when subject becomes indirect subject. Consider following examples.

Ali kitAbu likHiyO ----- (52a)
 Ali.M.Sg book.Nom.M.Sg write.PastPart.M.Sg
 Ali wrote a book.

kitAbu Ali=kHAN likhiyO vayO / likHijiyO (52b)
 book.Nom.M.Sg Ali-Agent write.Past.Pass.M.Sg
 Book was written by Ali.

mAu Ali=kHAN BArU-khE ----- (52c)
 mother.Nom Ali=Agent child.M.Sg-Dat

kHAdHO khArAyO
 food.Nom.M.Sg ate.caus1
 Mother caused Ali to feed food to the child.

In (52a) Ali is subject with nominative case and kitAbu is direct object. However, when this sentence is converted into passive voice as shown in (52b), the nominative subject becomes subject in agentive case marked by agentive case marker “kHAN”. Position of subject does not matter as Sindhi is partially free word order language.

Oblique arguments (see Section 2) are also marked by agentive case; in (52c) “mAu” is subject, “kHAdHO” is direct object, “BArU” is indirect object and “Ali” is oblique in agentive case. Keep in mind that oblique arguments are used with causative verbs (khAraAyO in this case).

3.2.11 Possessive / Genitive Case

Noun phrase in possessive case is marked by genitive case marker “jO”. “jO” is different than other case markers due to number and gender inflections. NP with possessive construction contains possessor noun as well as possessed noun. Consider following examples.

CHOKirla=jO kitAbu ----- (53a)
 girl.Obl=Gen.M.Sg book.Nom.M.Sg
 Girl's book.

CHOkirE=jUN boy.Obl=Gen.F.Pl Boy's Vehicles.	gADiUN Vehicles.Nom.F.Pl	----- (53b)
CHOkiriyan=jO Girls.Obl=Gen.M.Sg Girls' room.	kamrO room.Nom.M.Sg	----- (53c)

Sentence (53a) has a possessor “CHOkir” (girl) and possessed noun “kitAbu” (book), both are singular nouns, however possessor is feminine and possessed noun is masculine. Case marker must agree with possessed noun (which is head noun in fact) in number and gender. In (53b) this can be seen that possessor is masculine singular noun and possessed head noun is feminine plural which is agreeing with “jUN” (feminine plural postposition). In (53c) same phenomenon can be observed where possessed head noun is masculine singular, so is the case with postposition “jO”; while possessor noun “CHOkirani” is masculine plural. Possessor neither need to agree with case marker nor with possessed noun phrase. Genitive case can also be marked morphologically in first person and second person pronouns. As shown in following examples.

muhiNjO my.Gen.M.Sg My book.	kitAbu book.Nom.M.Sg	----- (54a)
tuhijA your.Gen.M.Pl Your apples.	sUfa apples.Nom.M.Pl	----- (54b)

In (54a) muhiNjO is equivalent morphological form of first person oblique pronoun “mUN” and singular masculine postposition “jO” (mUN-jO). Same is the case with second person “tUN” and plural masculine postposition “-jA” in (54b). Genitive case is therefore formed by genitive case marker and

sometimes with morphological genitive form in case of first and second person pronouns. Genitive case marking is different than other case markers as it implies number and gender agreement between head noun and postposition. NP in genitive case can act as subject, object or indirect object. This can be seen in following examples where “muhiNjO CHOkirO” in (55a) is a subject, “CHOkirE jO kitAbu” in (55b) is an object and “CHOkirE jE BilE” in (55c) is indirect object.

muhiNjO	CHOkirO	dORiyO	-----	(55a)
my.Gen.M.Sg	book.Nom.M.Sg	run.PastPart.M.Sg		
My boy ran.				

mUN	CHOkirE-jO	kitAbu	kHaNyO	-----	(55b)
I.Obl	boy.Obl-Gen.Nom.M.Sg	book.Nom.M.Sg	take.Pastpart.M.Sg		
I took boy's book.					

mUN	CHOkirE-jE	BilE-kHE	-----	(55c)
I.Obl	boy.Obl-Gen.Obl.M.Sg	cat.Obl.M.Sg-Dat		

kHIru	DinU
milk.Nom.M.Sg	give.PastPart.M.Sg
I gave milk to boy's cat.	

However, in LFG theory within NP the possessor noun is considered as specifier and is represented by SPEC instead of ADJUNCT in functional structure (see Section 4.5).

3.2.12 Case Marking with Case Phrase

Various grammatical cases discussed above are marked on noun phrases. A grammatical constituent case phrase is therefore required to cope with syntactic case marking on noun phrases. This case phrase is represented as KP instead of CP as CP is used for complement phrase in LFG notations. Formal definition of KP is given below.

KP → NP (K).

NP → { PPP* (ADJP) N | ADJP | PPrOn | NPPron (NC)} CPRel.

ADJP → ADJ (ADJ * CONJ ADJ).

PPP → N PP.

It can be seen here that in KP, NP is followed by a case marker (K); case marker can be any of the case markers discussed above excluding the genitive case marker. Moreover, K is optional as noun in NP can have morphological case. Genitive case marking is treated separately due to number and gender agreement between possessed NP and genitive case marker.

Possessive case phrase is defined separately and is called KPPoss; this KPPoss is usually a part of noun phrase and represents a noun phrase subject object or indirect object possessed by specifier (SPEC) NP. Formally this can be defined as given below.

KPPoss → NP KPoss;

NP → KPPoss NP.

Here KPoss is genitive case marker which follows NP. And definition here contains indirect recursion as KPPoss is defined in NP terms. This recursive definition covers complex genitive constructions. Following are examples of genitive constructions.

Ali=jO	kitAbu	----- (56a)
Ali=Gen.Nom.Sg.Masc	book.Nom.Sg.Masc	
Ali's book.		

Ali=jE	bHAu=jO	kitAbu ----- (56b)
Ali=Gen.Obl.Sg.Masc	brother=Gen.Nom.Sg.Masc	book.Nom.Sg.Masc
Ali's brother's book.		

Ali=jE	bHAu=jI	----- (56c)
Ali=Gen.Obl.Sg.Masc	brother=Gen.Nom.Sg.Fem	

dHla=jA kitAba
 daughter=Gen.Pl.Masc books.Nom.Pl.Masc
 Ali's brother's daughter's books.

(56a) is simple genitive construction where first phrase “ali jO” is defined by KPPoss where ali is NP and jO is KPoss followed by NP kitAbu. Derivation for simple genitive construction is given below and parse tree is shown in Figure VI-1 (a).

NP → KPPoss NP
 → NP KPoss NP
 → ali KPoss NP
 → ali jO NP
 → ali jO N
 → ali jO kitAbu

(56b) is bit more complicated and sentence generation from above rules can be seen below:

NP → KPPoss NP
 → NP KPoss NP
 → ali KPoss NP
 → ali jE NP
 → ali jE KPPoss NP
 → ali jE NP KPoss NP
 → ali jE bHAu KPoss NP
 → ali jE bHAu jO NP
 → ali jE bHAu jO kitAbu

Parse tree for above derivation is shown in Figure V-1(b). The possessive case phrase rule discussed above allows us to generate the incorrect parse as shown in Figure V-2. The valid parse is as shown in Figure V-1 (b). This

problem must be taken care of during implementation (see Chapter VI, section 3).

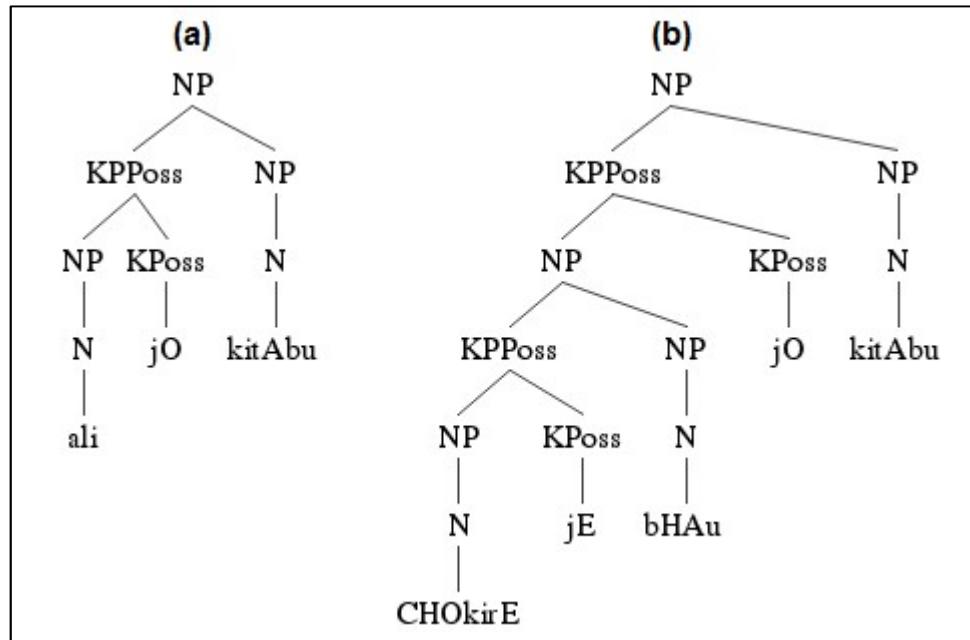


Figure III - 1: Possessive Noun Phrase Examples.

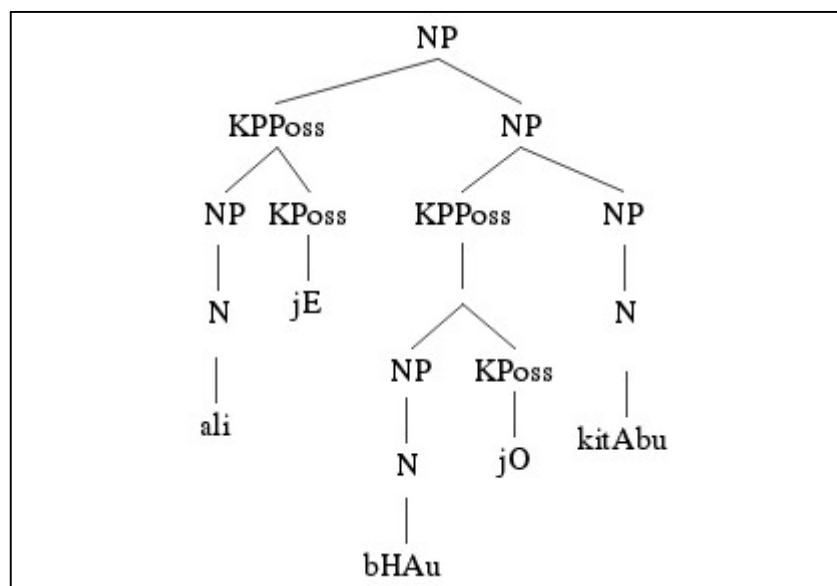


Figure III - 2: Incorrect Possessive Noun Phrase Parse.

In the same way (56c) can also be derived by using above rules.

3.3 Verb Sub Categorization

Sub categorization defines the argument structure of lexical items typically verbs. Verbs usually represent predicates which require one or more arguments (which co-occur with these predicates in a sentence/clause) for a sentence to be grammatical. In LFG theory predicates subcategorize for different grammatical functions which include: SUBJ (Subject), OBJ (Object), OBJ2 (Secondary Object), COMP (Complement), XCOMP (Infinitive Complement), and OBL (Oblique). As discussed in Chapter IV Sindhi verbs have morphological causative forms; these forms cause change in number and type of arguments in their verbal subcategorization frames. Following sections discuss Sindhi verbal subcategorization with respect to different grammatical functions listed above.

3.3.1 Subject

Like most of the linguistic theories LFG also assumes that verbs must subcategorize for subjects. Generally subject and verb are considered basic parts of a sentence. However, in Sindhi this is not always the case as there may be sentences without any subject or object. This subject condition is also challenged by researchers working on Hindi (Mohanan, 1994), Dutch (Zaenen, 1989) and German (Butt, et., al, 1999). In Sindhi subject condition is not necessary in various sentences containing passivized form of verbs; in these sentences subject is either not present or hidden in pronominal suffixes. In

following example (57a). the verb “likHu” (write) is subcategorized for subject where “Ali” is the subject; however, subject is suppressed due to passivization in (57b). and (57c). (passivization of intransitive verbs). (57d) shows the causative form of “dORi” with increase in number of arguments including subject and object.

Ali dORE thO ----- (57a)
 Ali run.Aoirst be.Sg.M.Pres.Aux

Ali runs.

(↑PRED)=’DORi<(↑SUBJ)>’

hitE dORijE thO ----- (57b)
 Here.Adv run.Pass.Cont be.Sg.M.Pres.Aux

Running takes place here.

(↑PRED)=’DORi<(NULL)>’

hitE dORibO AhE ----- (57c)
 Here.Adv run.Pass be.Cop.Sg

Running takes place here.

(↑PRED)=’DORi<(NULL)>’

Ali kutO dORAyO ----- (57d)
 Ali dog.Nom.M.Sg run.caus.M.Sg

Ali made dog run.

(↑PRED)=’DORi <(↑ SUBJ) (↑ OBJ)> ’

‘dORibO’ and ‘dORijE’ constructions are considered as passive constructions in this study as all the grammarians of Sindhi consider these constructions as passive. However, use of auxiliary and copula verbs suggest that these constructions can be some kind of participles; in that case, subcategorization will be with copula ‘be’ and main argument will be participle as given below (Note: passive constructions are considered in implementation instead of participle constructions).

(↑PRED) =’be<(dORijE)>’

(↑PRED) =’be<(dORibO)>’

Sometimes identification of subject in a clause/sentence in Sindhi is a

problem and standard tests like verb agreement, case morphology, reflexive pronoun bounds need to be tested and standard subject identification tests need to be defined. For example, consider following sentence where “CHOkirO” and “bHOlirO” both are candidates for subject as well as object. Their morphological case is same and due to free word order; they can replace each other’s position and can have equivalent order “bHOlirO CHOkirO mArE tHO”.

CHOkirO	bHOlirO	mArE	tHO ----- (58)
Boy.Nom	Monkey.Nom	Kill.Aorist	be.Sg.M.Pres.Aux
The boy kills the monkey.			

Thematic hierarchy defined by (Bresnan, 2001) and (Dalrymple, 2001) provides basis for identification of different grammatical functions. The hierarchy is shown below in Figure III-3. The “>” sign in hierarchy shows that left hand side role has higher prominence than the right-hand side (given below) role. Usually agent and experiencer roles define subjects, patient and theme roles define objects, goal and beneficiary roles define secondary or indirect objects, locative, instrument, source, and goal roles either define oblique arguments or adjuncts (Rizvi, 2007). However, argument identification problem discussed above remains unsolved and needs further investigation.

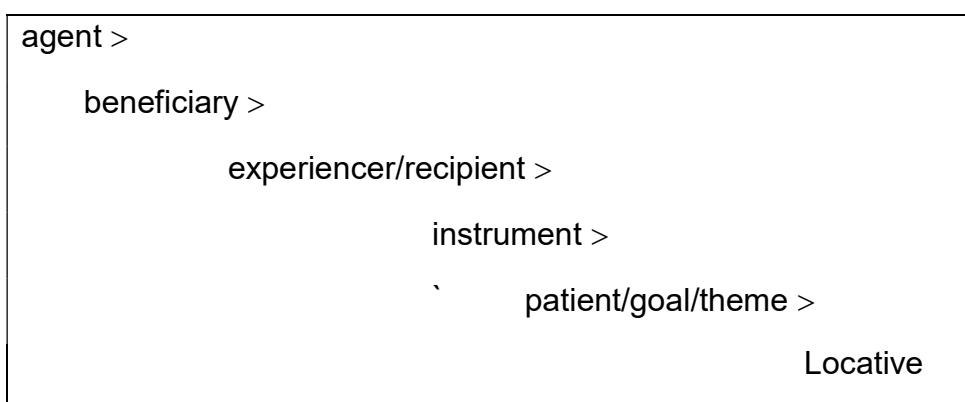


Figure III - 3: Thematic Hierarchy to Identify Grammatical Roles.

3.3.2 Object

Transitive verbs subcategorize for subjects and objects. Second argument is usually an object. In Sindhi transitive verbs are either original or derived from intransitive verbs by morphological inflection as shown in (57d). (59a – 59d) show examples of transitive verb “likHu” which subcategorizes for subject as well as object. Verbs always subcategorize for their arguments but some languages can show argument drop; Sindhi also drops arguments. This can be seen in transitive verbs where subject as well as object is not a necessary condition and alternatively “likHu” can subcategorize for only subject or suppressed object (subject) or even neither of these as shown in (59b – 59f).

Ali KitAbu likHE tHO ----- (59a)
 Ali.Nom.M.Sg book.Nom.M.Sg write.Aoirst.3P.Sg be.Sg.M.Pres.Aux
 Ali Writes a book.
 (↑PRED)=’LIKHU<(↑SUBJ) (↑OBJ)>

Ali likHE tHO ----- (59b)
 Ali.Nom.M.Sg write.Aoirst.3P.Sg be.Sg.M.Pres.Aux
 Ali Writes.
 (↑PRED)=’LIKHU<(↑SUBJ)>

KitAbu likHijE tHO ----- (59c)
 book.Nom.M.Sg write.Pass.Aorist.3P.Sg be.Sg.M.Pres.Aux
 Book is being written/Book writing takes place.
 (↑PRED)=’LIKHU<(NULL (↑SUB)>

KitAbu likHibO AhE ----- (59d)
 book.Nom.M.Sg write.Pass.Fut be.Cop.Sg
 Book writing takes place.
 (↑PRED)=’LIKHU<(NULL ↑SUB)>

likHibO AhE ----- (59e)
 write.Pass.Fut.Sg.Masc be.Cop.Sg
 Writing takes place.
 (↑PRED)=’LIKHU<(NULL)>

likHijE tHO ----- (59f)
 write.Pass.Aorist.Sg be.Aux.Pres.Sg.Masc
 (It's) being written.
 (↑PRED)=’LIKHU<(NULL)>’

3.3.3 OBJ2 (Secondary Objects)

Sindhi causatives are inflectional forms of intransitive and transitive verbs. Formation of causatives increases the valency of verb and causes changes in verbal sub categorization and semantics. (57a) and (57d) show such argument increasing examples of intransitive verbs. More complex causative formation of intransitive verbs is also possible and is discussed in ADJUNCT section (section 2.6). Consider transitive verb “kHAu” (usually transitive verbs subcategorize for both subject and object). However, the causative form “kHArAi” of “kHAu” will have a secondary object argument as well known as OBJ2. Secondary object in Sindhi is marked by postposition “kHE”. Example of such construction is shown in (60).

mAu BAru=kHE SUfa khArAyo ---- (60)
 mother.Nom.F.Sg child.M.Sg=Dat Apple.Nom.M.Sg eat.caus
 SUB OBJ2 OBJ
 Mother made child eat an apple
 (↑PRED)=’KhAu-cau<(↑SUB)(↑OBJ2)(↑OBJ)>’

3.3.4 Oblique

An oblique argument is neither subject nor direct or indirect object. Generally, it looks like an adjunct (discussed below) and like Urdu differentiating it from adjunct is bit complex due to presence of different case markers (Rizvi, 2007). Oblique markers are separated from postpositions and

possession markers. In Urdu oblique is marked by agentive case marker “sE”.

Same is the case with Sindhi where agentive case marker “kHAN” is used.

Consider following example where agentive oblique argument “plu” (father) is marked with “kHAN”. It may be noted that agentive argument must be an animate noun.

mAu	plu=kHAN	bAru	kHE	mAnl	kHArArAl	-----	(61)
Mother	Father=Obl	Child	To	Bread	Eat.caus2		
SUB	OBL	OBJ2	Dative	OBJ	Caused	Eating	
							Mother caused child to have bread by father.
							(↑ PRED)=’khAu<(↑ SUB) (↑ OBL) (↑ OBJ2) (↑ OBJ)>’

As compared to core arguments (SUBJ, OBJ, OBJ2) oblique arguments are less syntactic as they do not play any role in functional control constructions and they do not agree with verb or other clausal elements. Generally, oblique arguments are considered less strictly syntactic elements than the core elements (Falk, Y. N., 2000).

3.3.5 Complement and Open Complement

Arguments of a verb can be other clauses. In LFG theory two different types of clausal arguments are COMP (complement) and XCOMP (open complement). Complements usually complete the meaning of expression in a sentence. In open complement (XCOMP) subject is controlled from outside the complement clause and in close complement (COMP) subject is functionally controlled within the clause (Butt, et., al, 1999). Examples of COMP and XCOMP are shown in (62) and (63) respectively. In (62) it can be seen that COMP clause has its own subject “Ahmed” where as in (63) subject of XCOMP clause is referenced from main verb subject. Difference between open and closed complements is also expressed in terms of nonfinite and finite clauses

respectively.

Ali	sOchiyO	Ta	[Ahmed	kEIA	kHAE	tHO]	----- (62)
Ali	Thought	That	Ahmed	Bananas	Eat	be.Sg.M.Pres.Aux	
SUB	V	COMP	COMP-	COMP-	COMP	Aux	
		Maker	SUB	OBJ	-V		
Ali thought that Ahmed eats bananas.							
(↑ PRED)=’sOchi<(↑ SUB) (↑ COMP)							
COMP: ‘kHAu (↑ SUB) (↑ OBJ)>’							

Ali	kitAbu	paRhaNra	cHahiyO	----- (63)
Ali.Nom.M.Sg	book.Nom.M.Sg	read.inf	Want.PastPart.M.Sg	
SUB	XC-OBJ	XCOMP-V	V	
Ali wanted to read a book.				
(↑ PRED)=’cHahu<(↑ SUB) (↑ XCOMP)				
XCOMP: ‘paRhu (↑ SUB) (↑ OBJ)>’				
(↑ XCOMP SUBJ) = (↑ SUBJ)				

3.3.6 Adjunct

Adjuncts are postpositional and adverbial phrases which are not subcategorized for verb. Adjuncts appear along-with different verbal arguments discussed above and may also include embedded clauses. Adjuncts are expresses by pre/postpositional phrases, adverbial phrases and simple adverbs. An example of ADJUNCT with instrumental case is shown in (64a) where “cHamcHE-sAN” is an example of ADJUNCT phrase. There can be more than one ADJUNCT phrases in a sentence (64b) and (64c) show examples of such cases.

mAu	cHamcHE=sAN	BAru=kHE	mAnI	khArAI	----- (64a)
Mother	spoon.Obl.M.Sg=inst	Child=dat	Bread	eat.caus	
SUB	ADJUNCT-OBJ	OBJ2	OBJ	V	
Mother made child to have food with spoon.					
(↑ PRED)=’khAu<(↑ SUB) (↑ OBJ2) (↑ OBJ)>’					
ADJUNCT: sAN <(↑ OBJ)>					

bHOliRO	bAG=mEN	vaNra=tE	kHAE	tHO	----- (64b)
monkey	garden=loc	tree=loc	Eat	be.Sg.M.Pres.Aux	

SUB	ADJUNCT	ADJUNCT	V	Aux
Monlkey eats on a tree in the garden.				
bHOliRO	bAG-mEN	vaNra-tE	cHamcHE=sAN -----	(64c)
monkey	garden-loc	tree-loc	spoon=inst	
SUB	ADJUNCT	ADJUNCT	ADJUNCT	
kHAE	tHO			
Eat	be.Sg.M.Pres.Aux			
V	Aux			
Monkey eats with spoon on a tree in the garden.				

3.3.7 Open Adjunct (XAdjunct)

Like XCOMP, XADJUNCT or open adjunct is a clause whose subject is controlled from outside the clause. In Sindhi, such clauses are marked by conjunctive participle. For example, in (65) “kHAI” (after eating) is followed by XADJUNCT clause “gHaru vayO” but the subject of XADJUNCT clause verb “vayO” is referenced from verb “kHAI”.

CHOkirO	mAnI	kHAI	gHaru	vayO -----	(65)
Boy	bread/meal	eat.ConjPart	Home	Went	
SUB /	OBJ	V	XADJ-OBJ	XADJ-V	
XADJ-SUB					
The boy went home after eating/having eaten the meal.					

4. IMPLEMENTATION

Inflectional morphology of various word classes discussed in Chapter IV is implemented by incorporating the inflection rules in finite state models using Xerox LEXC (Lexicon Compiler) (Karttunen & Beesley, 1992). Different morphological paradigms of nouns, pronouns, adjectives, adverbs and verbs are represented in finite state transducer scripts by using LEXC syntax. These scripts are compiled and transducers are generated which represent Sindhi

morphological lexicon. Different syntactic constructions discussed in Chapter V are implemented in LFG within XLE environment by incorporating the compiled transducer through morphology syntax interface. Following sections discuss the implementation details of Sindhi Morphology and Syntax in LEXC and XLE. Section 1 discusses the implementation details of Sindhi morphology, section 2 discusses syntax implementation details in XLE along-with nominal syntax implementation, verbal syntax implementation is discussed in section 3.

4.1 Implementing Sindhi Morphology (Simple Word Classes)

Following sections discuss LEXC implementation details of different morphological classes of Sindhi. It may be noted that verb morphology and pronominal suffixation morphology are discussed in a separate section (Section 4.2).

4.1.1 Sindhi Noun Morphology

As discussed in Chapter IV Sindhi nouns are inflected by number gender and case. Different paradigms are modeled in FST rules and resulting transducers act as function machines in which either upper side represents the input and lower side represents the output or vice versa. The reversible property of these FSTs makes them very useful. When lower side is used for input these FSTs function as morphological analyzers and when upper side is used for input these will function as surface form generators. For example, if input on upper side is “CHOkir+Pl” the output at lower side will be “CHOkirA”.

Here “CHOkir” (boy) is stem word and “+Pl” is plural tag, when this information is given as input to transducer it generates the surface form “CHOkirA” (boys) plural of the stem word given as input. However, when this process is reversed i.e. when “CHOkirA” is input it will produce morphological analysis of the surface word i.e. “CHOkir+Pl”.

While defining finite state lexicon for Sindhi nouns in LEXC a Root lexicon named Nouns is defined which is further extended to various sub lexicons. These sub lexicons model various inflectional paradigms of nouns. LEXC fragment in Figure III-4 shows the root lexicon Nouns along-with N_Cat1 sub lexicon. Stem form of noun followed by sequence of tags can be seen here. This sequence represents various features of the noun. Stem along-with these features will produce intermediate word form shown after colon “:” following the tag sequence. This intermediate form is further inflected based on various feature sequences defined in sub lexicon N_Cat1. Consider the stem form and tag sequence given below:

CHOkir+Noun+Common+Count+Animate

This will produce intermediate animate common count noun form “CHOkir”, this transducer is followed by another transducer in series (via N_Cat1 sub lexicon link) which takes further input tags as shown below:

+Sg+Masc+Nominative

This produces the singular masculine nominative morpheme “O”. The overall concatenated tag sequence preceded by stem (upper side) and concatenated output (lower side) are given below.

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

Intermediate: CHOkir O

Lower: CHOkirO

```

!SINDHI NOUN MORPHOLOGY
!AUTHOR MUTEE U RAHMAN

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)
CHOkir0+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     #;
+Sg+Fem+Vocative:I        #;
+Sg+Fem+Oblique:I         #;
+Pl+Masc+Nominative:A     #;
+Pl+Masc+Oblique:ani      #;
+Pl+Masc+Vocative:aO      #;
+Pl+Fem+Nominative:yUN    #;
+Pl+Fem+Oblique:yani      #;
+Pl+Fem+Vocative:yUN      #;

```

Figure III - 4: Sindhi Noun Morphology LEXC Snapshot

Table III - 67: Different Morphological Forms of Common Noun Stem
“CHOkir”

Tag/Feature sequence	Intermediate Forms	Surface Form
CHOkir+Noun+Common+Count+Animate+Sg +Masc+Nominative	CHOKir O	CHOkirO
CHOkirO+Noun+Common+Count+Animate+Sg+Masc+Oblique	CHOKir E	CHOkirE
CHOkirO+Noun+Common+Count+Animate+Sg+Masc+Vocative	CHOKir A	CHOkirA
CHOkirO+Noun+Common+Count+Animate+Sg+Fem+Nominative	CHOKir Ia	CHOkirla
CHOkirO+Noun+Common+Count+Animate+Sg+Fem+Vocative	CHOKir I	CHOkirl
CHOkirO+Noun+Common+Count+Animate+Sg+Fem+Oblique	CHOKir I	CHOkirl
CHOkirO+Noun+Common+Count+Animate+Pl+Masc+Nominative	CHOKir A	CHOkirA
CHOkirO+Noun+Common+Count+Animate+Pl+Masc+Oblique	CHOKir ani	CHOkirani
CHOkirO+Noun+Common+Count+Animate+Pl+Masc+Vocative	CHOKir aO	CHOkiraO
CHOkirO+Noun+Common+Count+Animate+Pl+Fem+Nominative	CHOKir yUN	CHOkiryUN
CHOkirO+Noun+Common+Count+Animate+Pl+Fem+Oblique	CHOKir yani	CHOkiryani
CHOkirO+Noun+Common+Count+Animate+Pl+Fem+Vocative	CHOKir yUN	CHOkiryUN

While going from upper to lower side surface form “CHOkirO” of stem with features specified in tag sequence is generated; going from lower to upper will give following morphological analysis of noun “CHOkirO”.

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

This morphological analysis says that “CHOkirO” is a common animate count noun in singular masculine form with nominative case. In the same way, oblique morphological form (used as base for various syntactic cases of nouns) “CHOkirE” is generated by producing and concatenating the oblique morpheme “E” by input tag sequence “CHOkir+Noun+Common+Count+Animate+Sg+Masc+Oblique” and

output sequence “CHOkir” and “E”. All different morphological forms of stem “CHOkir” generated by above discussed transducer are shown in Table III-67. The table shows case (nominative and vocative), oblique form, number and gender inflections. Total twelve (12) different inflections of stem “CHOkir” are taken care of. A total of 21 different common noun categories are identified according to their inflectional properties. For every category, a different sub lexicon is defined. Usually proper nouns are not inflected therefore their entries only contain the feature tags. For example, the proper noun Pakistan has following entry in the lexicon.

Pakistan+Noun+Proper+Inanimate+Country+Masc+Sg+3rd:pAkist
An #;

```

LEXICON Root
Nouns;

LEXICON Nouns
.
.
.
!!!!!! PROPER NOUNS !!!!!!!
Ali+Noun+Proper+Animate+Name+Masc+Sg+3rd:ali    #;
dOd+Noun+Proper+Animate+Name+Mas+3rd:dOd      N_Cat1a;
Ahmed+Noun+Proper+Animate+Name+Masc+Sg+3rd:ahmed   #;
Saleem+Noun+Proper+Animate+Name+Masc+Sg+3rd:salIm   #;
Parveen+Noun+Proper+Animate+Name+Fem+Sg+3rd:parveen #;
.

LEXICON N_Cat1a
+Sg+Nominative:O      #;
+Sg+Oblique:E      #;
+Pl+Nominative:A      #;
+Pl+Oblique:ani      #;
+Sg+Vocative:A      #;

```

Figure III - 5: Proper Noun Entries in LEXC

It says that Pakistan is an inanimate masculine singular proper noun which is

a country and has surface form “pAkistAn”. Most of the proper nouns have this type of entry. However, in Sindhi there are exceptional cases of proper noun inflections. For example, a person name “dOdO” can have number, and case inflections “dOdA” (plural or vocative singular) and “dOdE” (oblique form). These inflections are handled by defining another sub lexicon shown in Figure III-5 along-with other proper noun entries.

4.1.2 Sindhi Adjective Morphology

Native Sindhi adjectives which belong to closed subclass are declined like nouns. Borrowed adjectives mostly do not decline. Adjective morphology is implemented in the same way as noun morphology is implemented. Few adjective entries are shown in Figure III-6 along-with sub lexicon entries. Entries include native adjective “suhiNrO” (beautiful) with number, gender and case declensions and borrowed adjective “khUbSUrat” (beautiful) with no declension. Adjective entries also include multi-character tag entries of different adjective subtypes like ordinal, cardinal and fractal. Morphological variations of adjective “suhiNrO” are identical to noun “CHOkirO”. However, the adjective “hika” (one) can have cardinal or multiplier inflections and can also have further inflections on gender, number and case like “suhiNrO”. Same is the case with “thOrO” which is further inflected by degree. The adjective “KHUbSUrat” is not inflected and remains in its stem form. All different morphological variations generated by Figure III-6 transducer are shown in Table III-68, Table III-69 and Table III-70.

```

!SINDHI ADJECTIVE MORPHOLOGY
!AUTHOR MUTEE U RAHMAN

Multichar_Symbols
+Adjective +Noun
+Ordinal
+Cardinal
+Fractal

+Fem +Masc !Gender
+Sg +Pl !Number

LEXICON Adjectives
suhinR0+Adjective:suhinR          AdjInf_Cat1;
suTH0+Adjective:suH               AdjInf_Cat1;
thOr0+Adjective:thOr             AdjInf_Cat2b;
hika+Adjective:hik              AdjInf_Cat2a;
KHUbSURat+Adjective:KHUbSURat #;

LEXICON AdjInf_Cat1
+Nominative+Masc+Sg:O           #;
+Oblique+Masc+Sg:E            #;
+Nominative+Masc+Pl:A          #;
+Oblique+Masc+Pl:ani          #;
+Nominative+Fem+Sg:I           #;
+Oblique+Fem+Sg:I             #;
+Nominative+Fem+Pl:iyUN         #;
+Oblique+Fem+Pl:iyani          #;

LEXICON AdjInf_Cat2a
+Cardinal:a                      #;
+Multiplier:ooNr                  MultiplierInflection;
0:aR                             AdjInf_Cat1;

LEXICON AdjInf_Cat2b
0:0                               AdjInf_Cat1;
0:aR                             AdjInf_Cat1;

LEXICON MultiplierInflection
0:a                             #;
0:0                             AdjInf_Cat1;

```

Figure III - 6: Adjective Morphology Entries

Table III - 68: Morphology of Native Adjective “suhiNrO” with Intermediate and Surface Forms

Tag/Feature sequence	Intermediate Forms	Surface Form
suhiNrO+Adjective+Nominative+Masc+Sg	suhiNr O	suhiNrO
suhiNrO+Adjective+Oblique+Masc+Sg	suhiNr E	suhiNrE
suhiNrO+Adjective+Nominative+Masc+Pl	suhiNr A	suhiNrA
suhiNrO+Adjective+Oblique+Masc+Pl	suhiNr ani	suhiNrani
suhiNrO+Adjective+Nominative+Fem+Sg	suhiNr I	suhiNrI
suhiNrO+Adjective+Oblique+Fem+Sg	suhiNr I	suhiNrI
suhiNrO+Adjective+Nominative+Fem+Pl	suhiNri yUN	suhiNriyUN
suhiNrO+Adjective+Oblique+Fem+Pl	suhiNri yani	suhiNriyani
suTHO+Adjective+Nominative+Masc+Sg	sutH O	sutHO
suTHO+Adjective+Oblique+Masc+Sg	sutH E	sutHE
suTHO+Adjective+Nominative+Masc+Pl	sutH A	sutHA
suTHO+Adjective+Oblique+Masc+Pl	sutH ani	sutHani
suTHO+Adjective+Nominative+Fem+Sg	sutH I	sutHI
suTHO+Adjective+Oblique+Fem+Sg	sutH I	sutHI
suTHO+Adjective+Nominative+Fem+Pl	sutHi yUN	sutHiyUN
suTHO+Adjective+Oblique+Fem+Pl	sutHi yani	sutHiyani

Table III - 69: Morphology of Native Adjective “thOrO” and Borrowed Adjective “KHUbsUrat” with Intermediate and Surface Forms

Tag/Feature sequence	Intermediate Forms	Surface Form
thOrO+Adjective+Nominative+Masc+Sg	thOr aR O	thOraRO
thOrO+Adjective+Oblique+Masc+Sg	thOr aR E	thOraRE
thOrO+Adjective+Nominative+Masc+Pl	thOr aR A	thOraRA
thOrO+Adjective+Oblique+Masc+Pl	thOr aR ani	thOraRani
thOrO+Adjective+Nominative+Fem+Sg	thOr aR I	thOraRI
thOrO+Adjective+Oblique+Fem+Sg	thOr aR I	thOraRI
thOrO+Adjective+Nominative+Fem+Pl	thOr aR iyUN	thOraRiyUN
thOrO+Adjective+Oblique+Fem+Pl	thOr aR yani	thOraRiyani
thOrO+Adjective+Nominative+Masc+Sg	thOr O	thOrO
thOrO+Adjective+Oblique+Masc+Sg	thOr E	thOrE
thOrO+Adjective+Nominative+Masc+Pl	thOr A	thOrA
thOrO+Adjective+Oblique+Masc+Pl	thOr ani	thOrani
thOrO+Adjective+Nominative+Fem+Sg	thOr I	thOrI
thOrO+Adjective+Oblique+Fem+Sg	thOr I	thOrI
thOrO+Adjective+Nominative+Fem+Pl	thOr iyUN	thOriyUN
thOrO+Adjective+Oblique+Fem+Pl	thOr iyani	thOriyani
KHUbsUrat+Adjective		KHUbSUrat

Table III - 70: Morphology of Adjective “hika” with Intermediate and Surface Forms

Tag/Feature sequence	Intermediate Forms	Surface Form
hika+Adjective+Cardinal	hik a	Hika
hika+Adjective+Multiplier	hik ooNra	hikooNra
hika+Adjective+Multiplier+Nominative +Masc+Sg	hik ooNrO	hikooNrO
hika+Adjective+Multiplier+Oblique +Masc+Sg	hik ooNrE	hikooNrE
hika+Adjective+Multiplier+Nominative +Masc+Pl	hik ooNrA	hikooNrA
hika+Adjective+Multiplier+Oblique +Masc+Pl	hik ooNrani	hikooNrani
hika+Adjective+Multiplier+Nominative +Fem+Sg	hik ooNrI	hikooNrI
hika+Adjective+Multiplier+Oblique +Fem+Sg	hik ooNrI	hikooNrI
hika+Adjective+Multiplier+Nominative +Fem+Pl	hik ooNriyUN	hikooNriyUN
hika+Adjective+Multiplier+Oblique +Fem+Pl	hik ooNriyani	hikooNriyani
hika+Adjective+Nominative+Masc+Sg	Hik aR O	hikaRO
hika+Adjective+Oblique+Masc+Sg	hik aR E	hikaRE
hika+Adjective+Nominative+Masc+Pl	hik aR A	hikaRA
hika+Adjective+Oblique+Masc+Pl	hik aR ani	hikaRani
hika+Adjective+Nominative+Fem+Sg	hik aR I	hikaRI
hika+Adjective+Oblique+Fem+Sg	hik aR I	hikaRI
hika+Adjective+Nominative+Fem+Pl	hik aR iyUN	hikaRiyUN
hika+Adjective+Oblique+Fem+Pl	Hik aR iyani	hikaRiyani

4.1.3 Sindhi Pronoun Morphology

Pronoun morphology covers all seven types of pronouns discussed in Section 1.1.1. Due to irregular morphology, most pronouns do not have inflectional entries. Figure III-7 shows snapshot of pronoun entries file. Multichar symbols are defined to tag different types and subtypes of pronouns like demonstrative, reflexive, relative, etc.

```

!SINDHI PRONOUN MORPHOLOGY
!AUTHOR MUTEE U RAHMAN

Multichar_Symbols

+Pron
+Demon +WhP +Refl !Noun Types
+Proximate +Remote +IndPron +Relative +Corelative
+What +Who +How +When +Where +Which +HowMuch +HowBig
+Nominative +Oblique +Genitive !Pronoun Cases
+GenSg +GenPl +GenMasc +GenFem +GenOblique

+Fem +Masc !Gender
+Sg +Pl !Number
+1P +2P +3P !Person

LEXICON Root
Pronouns;

LEXICON Pronouns

AUN+Pron+1P+Sg+Masc+Nominative:AUN      #;
AUN+Pron+1P+Sg+Fem+Nominative:AUN      #;

mUN+Pron+1P+Sg+Masc:mU                  PN_1PInf;
tUN+Pron+2P+Sg+Masc:t                  PN_2PInf;
mUN+Pron+1P+Sg+Fem:mU                  PN_1PInf;
tUN+Pron+2P+Sg+Fem:t                  PN_2PInf;

LEXICON PN_1PInf
+Oblique:N                      #;
+Genitive:hiN                   PN_GNInf;

LEXICON PN_2PInf
+Nominative:UN                  #;
+Oblique:O                      #;
+Genitive:uhin                  PN_GNInf;

LEXICON PN_GNInf
+GenSg+GenMasc:jO                #;
+GenPl+GenMasc:jA                #;
+GenSg+GenFem:jI                 #;
+GenPl+GenFem:jUN                #;
+GenSg+GenMasc+GenOblique:jE    #;
+GenPl+GenMasc+GenOblique:jini  #;
+GenSg+GenFem+GenOblique:ji     #;
+GenPl+GenFem+GenOblique:jiyani #;

```

Figure III - 7: Pronoun Morphology Entries in LEXC

Following entries are examples of simple entries without any inflection.

```
AUN+Pron+1P+Sg+Masc+Nominative:AUN      #;
```

```
AUN+Pron+1P+Sg+Fem+Nominative:AUN      #;
```

Here entry “AUN” is defined as first person personal singular nominative pronoun with masculine or feminine gender. However, there are examples of inflectional entries as well for instance consider following entry.

```
mUN+Pron+1P+Sg+Masc:mU          PN_1PInf;
```

In this entry, personal pronoun “mUN” is defined with singular masculine attributes with further inflectional entry PN_1PInf (pronoun 1st person inflection) with terminal oblique form or more genitive inflections with intermediate genitive infix and suffix inflection as shown below.

```
LEXICON PN_1PInf
+Oblique:N      #;
+Genitive:hiN      PN_GNInf;
```

PN_GNInf (personal noun genitive inflection) defines further inflections with genitive number and genitive gender tags. These genitive attributes need to be agreed with possessed noun at syntax level. Different morphological forms generated by Figure III-7 entries are shown in Table III-71 and Table III-72

Following example explains genitive tags in more detail.

```
mUN+Pron+1P+Sg+Fem+Genitive+GenSg+GenMasc+GenObli:
mUhINjE
```

The entry “mUN” with first person singular feminine has genitive tags GenSg, GenMasc, and GenOblique these tags represent genitive singular, genitive masculine and genitive oblique form/case respectively. This agreement must be handled at syntax level where the possessed noun of first person, singular pronoun must be in singular number with masculine gender in oblique form.

Table III - 71: Intermediate and Surface Forms of First Person Pronoun
Generated by Figure III-7 Entries

Tag/Feature sequence	Intermediate Forms	Surface Form
AUN+Pron+1P+Sg+Masc+Nominative		AUN
AUN+Pron+1P+Sg+Fem+Nominative		AUN
mUN+Pron+1P+Sg+Fem+Oblique	mU N	mUN
mUN+Pron+1P+Sg+Fem+Genitive +GenSg+GenMasc+GenOblique	mU hiN jE	mUhiNjE
mUN+Pron+1P+Sg+Fem+Genitive +GenSg+GenFem	mU hiN jl	mUhiNjl
mUN+Pron+1P+Sg+Fem+Genitive +GenSg+GenFem+GenOblique	mU hiN jl	mUhiNjl
mUN+Pron+1P+Sg+Fem+Genitive +GenSg+GenMasc	mU hiN jo	mUhiNjo
mUN+Pron+1P+Sg+Fem+Genitive +GenPl+GenFem	mU hiN jUN	mUhiNjUN
mUN+Pron+1P+Sg+Fem+Genitive +GenPl+GenMasc+GenOblique	mU hiN Jini	mUhiNjini
mUN+Pron+1P+Sg+Fem+Genitive +GenPl+GenFem+GenOblique	mU hiN Jiyani	mUhiNjiyani
mUN+Pron+1P+Sg+Fem+Genitive +GenPl+GenMasc	mU hiN ja	mUhiNja
mUN+Pron+1P+Sg+Masc+Oblique	mU hiN	mUN
mUN+Pron+1P+Sg+Masc+Genitive +GenSg+GenMasc+GenOblique	mU hiN jE	mUhiNjE
mUN+Pron+1P+Sg+Masc+Genitive +GenSg+GenFem	mU hiN jl	mUhiNjl
mUN+Pron+1P+Sg+Masc+Genitive +GenSg+GenFem+GenOblique	mU hiN jl	mUhiNjl
mUN+Pron+1P+Sg+Masc+Genitive +GenSg+GenMasc	mU hiN jo	mUhiNjo
mUN+Pron+1P+Sg+Masc+Genitive +GenPl+GenFem	mU hiN jUN	mUhiNjUN
mUN+Pron+1P+Sg+Masc+Genitive +GenPl+GenMasc+GenOblique	mU hiN Jinni	mUhiNjini
mUN+Pron+1P+Sg+Masc+Genitive +GenPl+GenFem+GenOblique	mU hiN Jiyani	mUhiNjiyani
mUN+Pron+1P+Sg+Masc+Genitive +GenPl+GenMasc	mU hiN ja	mUhiNja

Table III - 72: Intermediate and Surface Forms of Second Person Pronoun Generated by Figure III-7 Entries

Tag/Feature sequence	Intermediate Forms	Surface Form
tUN+Pron+2P+Sg+Fem+Nominative	tUN	tUN
tUN+Pron+2P+Sg+Fem+Oblique	tO	tO
tUN+Pron+2P+Sg+Fem+Genitive+GenSg +GenMasc+GenOblique	Tu hiN jE	tuhiNjE
tUN+Pron+2P+Sg+Fem+Genitive+GenSg +GenFem	Tu hiN jl	tuhiNjl
tUN+Pron+2P+Sg+Fem+Genitive+GenSg +GenFem+GenOblique	Tu hiN jl	tuhiNjl
tUN+Pron+2P+Sg+Fem+Genitive+GenSg +GenMasc	Tu hiN jO	tuhiNjO
tUN+Pron+2P+Sg+Fem+Genitive+GenPI +GenFem	Tu hiN jUN	tuhiNjUN
tUN+Pron+2P+Sg+Fem+Genitive+GenPI +GenMasc+GenOblique	Tu hiN jini	tuhiNjini
tUN+Pron+2P+Sg+Fem+Genitive+GenPI +GenFem+GenOblique	Tu hiN jiyani	tuhiNjiyani
tUN+Pron+2P+Sg+Fem+Genitive+GenPI +GenMasc	Tu hiN jA	tuhiNjA
tUN+Pron+2P+Sg+Masc+Nominative	tUN	tUN
tUN+Pron+2P+Sg+Masc+Oblique	tO	tO
tUN+Pron+2P+Sg+Masc+Genitive+GenSg +GenMasc+GenOblique	Tu hiN jE	tuhiNjE
tUN+Pron+2P+Sg+Masc+Genitive+GenSg +GenFem	Tu hiN jl	tuhiNjl
tUN+Pron+2P+Sg+Masc+Genitive+GenSg +GenFem+GenOblique	Tu hiN jl	tuhiNjl
tUN+Pron+2P+Sg+Masc+Genitive+GenSg +GenMasc	Tu hiN jO	tuhiNjO
tUN+Pron+2P+Sg+Masc+Genitive+GenPI +GenFem	Tu hiN jUN	tuhiNjUN
tUN+Pron+2P+Sg+Masc+Genitive+GenPI +GenMasc+GenOblique	Tu hiN jini	tuhiNjini
tUN+Pron+2P+Sg+Masc+Genitive+GenPI +GenFem+GenOblique	Tu hiN jiyani	tuhiNjiyani
tUN+Pron+2P+Sg+Masc+Genitive+GenPI +GenMasc	Tu hiN jA	tuhiNjA

```

...
asIN+Pron+1P+Pl+Masc:as PN_Cat1;
hIa+Pron+3P+Sg+Fem+Nominative:hIa #;
hIa+Demon+Proximate+Sg+Fem+Nominative:hIa #;
hUa+Demon+Remote+Sg+Fem+Nominative:hUa #;
kHud+Pron+Ref1+Sg+Masc+Nominative:kHud #;
kHud+Pron+Ref1+Sg+Fem+Nominative:kHud #;
pahiNjO+Pron+Ref1+Genitive:pahiNj
GenitiveMFSGPL;

ina+Pron+3P+Sg+Masc+Oblique:ina #;
kO+IndPron+Nominative+Masc+Sg:kO #;

jehRO+Relative:jehR Wh_Inf2;
tehRO+Corelative:tehR Wh_Inf2;

LEXICON PN_Cat1

+Nominative:IN #;
+Oblique:AN #;

LEXICON GenitiveMFSGPL

+Sg+Masc:O #;
+Pl+Masc:A #;
+Sg+Fem:Ia #;
+Pl+Fem:UN #;
+Sg+Masc+Oblique:E #;
+Sg+Fem+Oblique:I #;
+Pl+Masc+Oblique:ani #;
+Pl+Fem+Oblique:iyani #;

```

Figure III - 8: Different Pronoun Types in LEXC Morphology

Section 3.2 explains this genitive case marking at syntax level in detail. Figure III-8 shows more entries of different types of pronouns including personal, demonstrative, reflexive, relative, correlative, and indefinite pronouns.

4.1.4 Sindhi Postposition Morphology

Postpositions in Sindhi are closed word class and mostly remain uninflected. However, few number gender and case inflections are possible. Postposition morphology is represented by LFG full form lexicon entries and is not implemented in XFST LEXC. Full form lexicon used to implement postposition morphology is shown in Figure III-9. Different postpositions like “tE”, “mEN” and “sAN” shown in first three entries do not inflect and their entries just define postposition form and case attributes. The entries “tE” and “mEN” are used in locative case; however, the postposition “sAN” may be used in two different cases i.e. when used with pronoun or animate noun the case will be participant and when used with an inanimate noun the case will be instrumental. The postposition “khE” is used as a case marker for accusative and dative cases. “khAN” is ablative inflection of “khE”. The postposition “jO” is a possessive marker and is dealt with a different entry KPoss due to its different agreement properties. “jO” has different inflections and every inflection is represented as a separate entry. First entry “jA” in following example represents a postposition form ‘of’ with cnum (case number) and cgend (case gender) and case attributes with plural, masculine and genitive values respectively. Consider following examples to understand the use of these attributes.

- (1) CHOkirE jO kitAbu AhE
- (2) CHOkirE jA kitAba Ahen

in first sentence the noun “CHOkirO” is in oblique form with genitive case which is marked by “jO” postposition.

RAHMAN SINDHI LEXICON (1.0)

tE	PP * (^ PP-FORM) = on (^ CASE)=loc.
mEN	PP * (^ PP-FORM) = in (^ CASE)=loc.
sAN	PP * (^ PP-FORM)=with { { (^ NSEM N-CONCEPT)=c animate (^ NTYPE NSYN)=c pronoun } (^ CASE)=part (^ NSEM N-CONCEPT)=c inanimate (^ CASE)=inst}.
jO	KPoss * (^ PP-FORM)=of (^ CNUM)=sg (^ CGEND)=masc (^ CASE)=gen.
jA	KPoss * (^ PP-FORM)=of (^ CNUM)=pl (^ CGEND)=masc (^ CASE)=gen.
jUN	KPoss * (^ PP-FORM)=of (^ CNUM)=pl (^ CGEND)=fem (^ CASE)=gen.
jI	KPoss * (^ PP-FORM)=of (^ CNUM)=sg (^ CGEND)=fem (^ CASE)=gen.
jE	KPoss * (^ PP-FORM)=of (^ CNUM)=sg (^ CGEND)=masc " (^ CN-FORM)=obl" (^ CASE)=gen.
khE	PP * (^ PP-FORM)=khe { (^ CASE)=acc (^ CASE)= dat}.
khAN	PP * (^ PP-FORM)=khan (^ CASE)=abl.

Figure III - 9: Full Form Lexicon Entries for Postpositons

The possessive phrase “CHOkirE jO” is analyzed as a noun “CHOkirO” with genitive case, singular CNUM, and masculine CGEND values. These attributes need to be agreed with PREDLINK “kitAbu” at syntax level. This agreement can also be seen in second sentence where plural PREDLINK “kitAba” is there because of plural case marker postposition “jA”. See Section 3.2.11 for more details.

4.1.5 Sindhi Adverb Morphology

Adverbs in Sindhi are rarely inflected, only when other classes (nouns, adjectives and few pronouns) are used as adverbs. Other classes when used as adverbs hold their inflectional properties and forms. However, words which belong only adverbs class are not inflected. Adverb morphology can either be implemented as a full form lexicon as postposition morphology discussed above or separate LEXC entries in FST definitions. Adverbs are implemented as straight entries in LEXC and compiled as finite state transducers. Figure III-10 shows a snapshot of LEXC entries file in which different adverb entries are given.

```

!Sindhi Adverbs Lexicon
!AUTHOR MUTEE U RAHMAN

Multichar_Symbols
+Adverb
+Sg +Pl
+Masc +Fem
+Manner +Space +Temp +Pron +Neg +Quantity +Affirm +N
+Adj +na +kona +nahe +konhe

LEXICON Root
Adverbs;

LEXICON Adverbs

ZarUr+Adverb:ZarUr          #;
ZAhir+Adverb+Adj:ZAhir       #;
mathE+Adverb+Space:mathE     #;
mathAN+Adverb+Space:mathAN   #;
aGyAN+Adverb+Space:aGyAN    #;
pOyAN+Adverb+Space:pOyAN    #;
hETHa+Adverb+Space:hETHa    #;
hETHAN+Adverb+Space:hETHAN  #;
kOna+Adverb+Neg+kona:kOna   #;
kOnhE+Adverb+Neg+konhe:kOnhE #;
nAhE+Adverb+Neg+nahe:nAhE   #;
na+Adverb+Neg+na:na         #;
ADO+Adverb+Space:ADO         #;
aGEI+Adverb+Temp:aGEI        #;
kaDNhN+Adverb+Temp:kaDNhN   #;
aJi+Adverb+Temp:aJi          #;
DANhN+Adverb+Space:DANhN    #;
tORE+Adverb:tORE             #;
bHale+Adverb+Affirm:bHale   #;
tAIN+Adverb+Temp:tAIN        #;
namUnE+Adverb:namUnE         #

```

Figure III - 10: Adverbs Lexicon Entries in LEXC

4.1.6 Sindhi Conjunction and Interjection Morphology

Conjunctions are closed word class and do not inflect. Subordinate conjunctions and coordinate conjunctions are represented as a full form lexicon in LFG. Figure III-11 shows snapshot of full form lexicon conjunction entries in LFG. Coordinate conjunctions have CONJ category and subordinate conjunctions have C category. Subordinate conjunction ta can also be used as intensifier (e.g. acha ta). Interjections are neither marked by number gender and case nor inflected; may have irregular inflections which are new interjections and are dealt as separate interjections. Interjections are also handled by using full form lexical entries in LFG.

RAHMAN SINDHI LEXICON (1.0)	
ain	CONJ * (^ COORD-FORM) = and.
nah	CONJ * (^ COORD-FORM)=nah.
tORE	CONJ * (^ COORD-FORM)=and.
yA	CONJ * (^ CORD-FORM)= or.
ta	C * (^ C-TYPE)=that;
	INTC * (^ INTCTYPE)=ta.

Figure III - 11: Full Form Lexicon Entries for Conjunctions

4.2 Implementing Sindhi Morphology (Verbs and Pronominal Suffixes)

Verb in Sindhi is a morphologically complex word class. As discussed in Section 2.9 verbs are marked by number, gender, case, tense, aspect and mood. Various categories of auxiliary verbs are also inflected by number,

gender, and case; auxiliaries may also be used as tense and aspect markers with inflections. Copula verbs also undergo morphological changes. Due to many different categories of verbs reasonably good number of tags is used while implementing verb morphology Figure III-12 shows the snapshot of multichar_symbol entries used while implementing verb morphology in LEXC. First three entries are used to tag verb, auxiliary and copula respectively. These tags categorize and tag the verbs. “+Contra” tag is used to tag auxiliary verb which is used as a marker of contrafactual mood. “+AhE”, “+thO”, “+payO”, and “+chukO” tags represent different forms of auxiliary verbs. +Comp tags the completion marker which plays important role in aspect identification at syntax level. “+Adjective” tag is also used to tag some verbal forms (verbal nouns) which are used as adjectives. Imperative, imperative polite, desiderative and infinitive verb forms are tagged by “+Imp”, “+Polite”, “+Des”, and “+Inf” respectively. As discussed earlier Sindhi has different causative formations four causative formations are modeled here and are tagged with “+Caus1”, “+Caus2”, “+Caus3”, and “+Caus4” respectively. Present participle, past participle, future participle, verbal noun and conjunctive participle form tags are also there followed by different tense forms which include “+Aorist”, “+Present”, “+Past”, “+PastInd”, and “+Future” tags. Active and passive forms are also tagged with “+Active” and “+Passive” tags. “+PsX” tag is used to mark pronominal suffixes. Different other tags used to represent pronominal suffixes are preceded by “+O” or “+S” prefix which represent object and subject attribute respectively. For example, “+O2P” is tag for second person object and “+SMasc” represents masculine subject.

Auxiliaries which are used as continuity markers are tagged with “+Cont” tag.

Rest of the tags are like other class tags discussed in previous sections.

Implementation of auxiliary verb, copula verb and main verb morphology is discussed in subsequent sections in detail.

```
! Sindhi Verbs Lexicon
! AUTHOR MUTEE U RAHMAN
Multichar_Symbols
+Verb +Aux +Cop +Contra +AhE +thO +payO +Comp +chuko
+Adjective
+Imp
+Rude +Polite +Des
+Inf
+Caus1 +Caus2 +Caus3 +Caus4
+PresPart +PastPart +FutPart +VerbNoun +ConjPart
+Aorist +Present +Past +PastInd +Future
+Passive +Active
+Psx
+OSg +OPl +SSg +SPl +S1P +S2P +O2P +O3P +OMF +SMF
+OMasc +SMasc +SObl +SFem +OFem
+Cont
+Sg +Pl
+Masc +Fem
+1P +2P +3P
```

Figure III - 12: Multi_Char Symbols Used in Verb Morphology

4.2.1 Auxiliary and Copula Verbs

Auxiliaries in Sindhi are marked by number, gender, person, and tense form. Depending on their syntactic role some auxiliaries may not be marked by some of these attributes. Consider the auxiliary verb “AhE” (is) in Sindhi and its finite state morphology shown in Figure III-13. In ‘auxiliaries’ lexicon entries “AhE” has five different entries which define various morphological

variations of “AhE”. The first entry AhE+Aux+AhE:Ah AuxInfCat1; defines the morphology of different aorist forms of “AhE” with number and person marking. Next three entries given below define the passive form of “AhE”, present tense singular and present tense plural forms respectively.

```
AhE+Aux+Sg+Passive+AhE:hujE      #; !passive form of Ahe
AhE+Aux+Sg+Present+AhE:AhE      #;
AhE+Aux+Pl+Present+AhE:Ahen      #;
```

Next entry given below defines the future and past tense forms of “AhE” by using intermediate lexicon AuxInfCat1a.

```
AhE+Aux+AhE:0          AuxInfCat1a;
LEXICON                AuxInfCat1a
0:0                   AuxFutureInf;
0:0                   AuxPastInf;
```

Various auxiliary forms generated by FST defined in Figure III-13 are shown in Table III-73.

```

LEXICON Root
Verbs;
Auxiliaries;

LEXICON Auxiliaries
AhE+Aux+AhE:Ah           AuxInfCat1; !Auxiliary Inflection Category 1
AhE+Aux+Sg+Passive+AhE:huje      #; !passive form of Ahe
AhE+Aux+Sg+Present+AhE:AhE      #;
AhE+Aux+Pl+Present+AhE:Ahen     #;
AhE+Aux+AhE:0                 AuxInfCat1a;
AhE+Cop:Ah                   AuxInfCat1;

LEXICON AuxInfCat1
+Aorist+Sg+1P:yAN          #;
+Aorist+Pl+1P:yUN          #;
+Aorist+Sg+2P:IN           #;
+Aorist+Pl+2P:yO           #;
+Aorist+Sg+3P:E            #;
+Aorist+Pl+3P:en           #;

LEXICON AuxInfCat1a
0:0                         AuxFutureInf;
0:0                         AuxPastInf;

LEXICON AuxFutureInf
+Future+Masc+Sg+1P:hUNDasi    #;
+Future+Masc+Pl+1P:hUNDAsIN   #;
+Future+Fem+Sg+1P:hUNDiyasi   #;
+Future+Fem+Pl+1P:hUNDiyUNSIN #;
+Future+Masc+Sg+2P:hUNDEN    #;
+Future+Masc+Pl+2P:hUNDa     #;
+Future+Fem+Sg+2P:hUNDiIN    #;
+Future+Fem+Pl+2P:hUNDyUN    #;
+Future+Masc+Sg+3P:hUND0     #;
+Future+Masc+Pl+3P:hUNDa     #;
+Future+Fem+Sg+3P:hUNDI     #;
+Future+Fem+Pl+3P:hUNDyUN    #;

LEXICON AuxPastInf
+Past+Masc+Sg+1P:huasi       #;
+Past+Masc+Pl+1P:huAsIN      #;
+Past+Fem+Sg+1P:huyasi       #;
+Past+Fem+Pl+1P:huyUNSIN    #;
+Past+Masc+Sg+2P:huIN        #;
+Past+Masc+Pl+2P:huA         #;
+Past+Fem+Sg+2P:huyIN        #;
+Past+Fem+Pl+2P:huyUN        #;
+Past+Masc+Sg+3P:huO         #;
+Past+Masc+Pl+3P:huA         #;
+Past+Fem+Sg+3P:hui          #;
+Past+Fem+Pl+3P:huyUN        #;

```

Figure III - 13: Auxiliary AhE Morphology in LEXC

Table III - 73: Different Forms of Auxiliary “AhE” Generated by Figure VI-10 Transducer

Tag/Feature sequence	Intermediate Forms	Surface Form
AhE+Aux+AhE+Aorist+Sg+1P	Ah yAN	AhyAN
AhE+Aux+AhE+Aorist+Pl+1P	Ah yUN	AhyUN
AhE+Aux+AhE+Aorist+Pl+2P	Ah yO	AhyO
AhE+Aux+AhE+Aorist+Sg+2P	Ah IN	AhIN
AhE+Aux+AhE+Aorist+Sg+3P	Ah E	AhE
AhE+Aux+AhE+Aorist+Pl+3P	Ah en	Ahen
AhE+Aux+Sg+Present+AhE	AhE	AhE
AhE+Aux+Pl+Present+AhE	Ahen	Ahen
AhE+Aux+Sg+Passive+AhE	HujE	hujE
AhE+Aux+AhE+Past+Masc+Sg+1P	Huasi	Huasi
AhE+Aux+AhE+Past+Masc+Pl+1P	huAsIN	huAsIN
AhE+Aux+AhE+Past+Masc+Pl+3P	huA	huA
AhE+Aux+AhE+Past+Masc+Pl+2P	huA	huA
AhE+Aux+AhE+Past+Masc+Sg+2P	huiN	huiN
AhE+Aux+AhE+Past+Masc+Sg+3P	huO	huO
AhE+Aux+AhE+Past+Fem+Sg+1P	Huyasi	Huyasi
AhE+Aux+AhE+Past+Fem+Sg+2P	huyIN	huyIN
AhE+Aux+AhE+Past+Fem+Pl+1P	huyUNsIN	huyUNsIN
AhE+Aux+AhE+Past+Fem+Pl+2P	huyUN	huyUN
AhE+Aux+AhE+Past+Fem+Pl+3P	huyUN	huyUN
AhE+Aux+AhE+Past+Fem+Sg+3P	hul	hul
AhE+Aux+AhE+Future+Masc+Sg+1P	hUNDasi	hUNDasi
AhE+Aux+AhE+Future+Masc+Sg+2P	hUNDEN	hUNDEN
AhE+Aux+AhE+Future+Masc+Sg+3P	hUND O	hUND O
AhE+Aux+AhE+Future+Masc+Pl+1P	hUNDAsIN	hUNDAsIN
AhE+Aux+AhE+Future+Masc+Pl+2P	hUNDa	hUNDa
AhE+Aux+AhE+Future+Masc+Pl+3P	hUNDa	hUNDa
AhE+Aux+AhE+Future+Fem+Sg+1P	hUNDiyasi	hUNDiyasi
AhE+Aux+AhE+Future+Fem+Sg+2P	hUNDilN	hUNDilN
AhE+Aux+AhE+Future+Fem+Sg+3P	hUNDl	hUNDl
AhE+Aux+AhE+Future+Fem+Pl+1P	hUNDiyUNsIN	hUNDiyUNsIN
AhE+Aux+AhE+Future+Fem+Pl+2P	hUNDyUN	hUNDyUN
AhE+Aux+AhE+Future+Fem+Pl+3P	hUNDyUN	hUNDyUN

“AhE” can also be used as copula verb. Morphology of copula “AhE” is identical to auxiliary “AhE” only “+Aux” tag is replaced by “+Cop” and auxiliary form tag “+AhE” is not used.

```

LEXICON Root
Verbs;
Auxiliaries;

LEXICON Auxiliaries

thO+Aux+thO:th          AuxInfCat4;
payO+Aux+payO+Cont:pa    AuxInfCat3;

LEXICON AuxInfCat3

0:y                      AuxInfCat4;
0:i                      PassiveEnding;

LEXICON AuxInfCat4

+Present+Sg+Masc:O       #;
+Present+Pl+Masc:A       #;
+Present+Sg+Fem:I        #;
+ConjPart:I               #;
+Present+Pl+Fem:yUN      #;

LEXICON PassiveEnding

+Passive+Sg+Masc:jE      #;
+Passive+Sg+Fem:jE        #;
+Passive+Pl+Masc:jani     #;
+Passive+Pl+Fem:jani      #;

+Passive+Sg+Masc+ConPart:jI   #;
+Passive+Sg+Fem+ConPart:jI  #;
+Passive+Pl+Masc+ConPart:jI  #;
+Passive+Pl+Fem+ConPart:jI   #;

+Passive+Sg+Masc:bO      #;
+Passive+Sg+Fem:bI        #;
+Passive+Pl+Masc:bA        #;
+Passive+Pl+Fem:biyUN     #;

```

Figure III - 14: Morphology of Auxiliaries “thO” and payO in LEXC

Auxiliary “tHO” and “payO” are used as present tense marker and present tense continuity marker respectively. Figure III-14 shows the snapshot of LEXC implementation of these auxiliaries. “tHO” can have gender and number inflections in present tense (as tHO is a present tense marker). However conjunctive participle inflection is also defined here for “tHO”. “payO” is continuity marker in present tense and is inflected in number, gender and passive forms. Various forms of “tHO” and “payO” generated by FST of Figure III-14 are shown in Table III-74. Eight inflectional categories of auxiliary verbs are implemented which define the morphology of auxiliary and copula verbs.

Table III - 74: Intermediate and Surface Forms of “tHO” and “payO”

Tag/Feature sequence	Intermediate Forms	Surface Form
thO+Aux+thO+Present+Sg+Masc	th O	thO
thO+Aux+thO+Present+Pl+Masc	th A	thA
thO+Aux+thO+Present+Sg+Fem	th I	thI
thO+Aux+thO+ConjPart	th I	thI
thO+Aux+thO+Present+Pl+Fem	th yUN	thyUN
payO+Aux+payO+Cont+Present+Sg+Masc	pa y O	payO
payO+Aux+payO+Cont+Present+Pl+Masc	pa y A	payA
payO+Aux+payO+Cont+Present+Sg+Fem	pa y I	payI
payO+Aux+payO+Cont+ConjPart	pa y I	payI
payO+Aux+payO+Cont+Present+Pl+Fem	pa y yUN	payyUN
payO+Aux+payO+Cont+Passive+Sg+Fem	pa i jE	paijE
payO+Aux+payO+Cont+Passive+Sg+Masc	pa i jE	paijE
payO+Aux+payO+Cont+Passive+Pl+Masc	pa i jani	Paijani
payO+Aux+payO+Cont+Passive+Pl+Fem	pa i jani	Paijani
payO+Aux+payO+Cont+Passive+Sg+Masc +ConPart	pa i jl	paijl
payO+Aux+payO+Cont+Passive+Sg+Fem+ConPart	pa i jl	paijl
payO+Aux+payO+Cont+Passive+Pl+Masc+ConPart	pa i jl	paijl
payO+Aux+payO+Cont+Passive+Pl+Fem+ConPart	pa i jl	paijl
payO+Aux+payO+Cont+Passive+Sg+Masc	pa i bO	paibO
payO+Aux+payO+Cont+Passive+Sg+Fem	pa i bl	paibl
payO+Aux+payO+Cont+Passive+Pl+Masc	pa i bA	paibA
payO+Aux+payO+Cont+Passive+Pl+Fem	pa i biyUN	paibiyUN

4.2.2 Verbs

Verb morphology is implemented on same patterns like other morphologies discussed above. Different morphological classes are defined by sub-lexicon definitions. Morphological implementation of transitive verb “likHu” (write) and intransitive verb “dOri” (run) is discussed below in detail. Figure III-15 shows the lexicon entries of these two verbs. Sub lexicon definition verbstem1 is shown in Figure III-16 which defines the transitive verb morphology further.

LEXICON Verbs	
likHu+Verb:likH	VerbStem1;!Transitive Likhu=Write
dORi+Verb:dOR	VerbStem2;!Intransitive Type-1

Figure III - 15: Lexicon Entries of Transitive Verb “likHu” and Intransitive Verb “dOri” in LEXC

Further inflections in verbstem1 lexicon include infinitive, pronominal suffixes and passive verb forms. Pronominal suffixes are discussed in next section. Infinitive and passive ending sub-lexicon are shown in Figure III-17. Morphological implementation of verbs shown in Figure III-9, Figure III-16 and Figure III-17 covers various verbal forms including infinitive, imperative, causative (four types), polite imperative, desiderative, present participle, past participle, future participle, conjunctive participle, verbal noun, aorist, future and passive along-with number and gender inflections.

```

LEXICON VerbStem1
0:a                                Infinitive;
+Psx:iy                                PSuffix1;
+Psx+PastPart+Sg:iyO                  PSuffix2;
+Psx+PastPart+Pl:iyA                  PSuffix2;
0:i                                PassiveEnding;
+Imp+Caus1+Sg+2P:u                      #;
+Imp+Caus1+Pl+2P:O                      #;
+Imp+Polite+Caus1+Sg+2P:iji            #;
+Imp+Polite+Caus1+Pl+2P:ijo             #;
+Imp+Des+Caus1+Sg+2P:ijANI            #;
+Imp+Des+Caus1+Pl+2P:ijaO              #;
+PresPart+Masc+Sg:andO                #;
+PresPart+Masc+Pl:andA                #;
+PresPart+Fem+Sg:andI                #;
+PresPart+Fem+Pl:andyUN              #;
+PastPart+Masc+Sg:yO                 #;
+PastPart+Masc+Pl:yA                 #;
+PastPart+Fem+Sg:I                   #;
+PastPart+Fem+Pl:yUN                 #;
+PastPart:yala                         #;
+FutPart:aNrU                          #;
+FutPart+Sg:aNrU                        #;
+FutPart+Pl:aNrA                        #;
+VerbNoun:andaRa                       #;
+ConjPart:I                            #;
+Imp+Caus2+2P:Ai                        #;
+Imp+Caus3+2P:ArAi                      #;
+Imp+Caus4+2P:ArArAi                    #;
+Aorist+Sg+1P:AN                         #;
+Aorist+Pl+1P:UN                         #;
+Aorist+Sg+2P:IN                         #;
+Aorist+Pl+2P:O                          #;
+Aorist+Sg+3P:E                          #;
+Aorist+Pl+3P:ani                         #;
+Future+Sg+Masc+1P:andasu               #;
+Future+Sg+Fem+1P:andiyasu              #;
+Future+Pl+Masc+1P:andAsIN              #;
+Future+Pl+Fem+1P:andyUNsIN              #;
+Future+Sg+Masc+2P:andIN                #;
+Future+Sg+Fem+2P:andiIN                #;
+Future+Pl+Masc+2P:andaO                #;
+Future+Pl+Fem+2P:andiaO                #;
+Future+Sg+Masc+3P:andO                 #; !Present Participle form
+Future+Sg+Fem+3P:andI                 #; !Present Participle form
+Future+Pl+Masc+3P:andA                 #; !Present Participle form
+Future+Pl+Fem+3P:andyUN                #; !Present Participle form

```

Figure III - 16: Verbstem1 Sublexicon Definition

LEXICON Infinitive	
Formation	
+Inf:Nra	#;
LEXICON PassiveEnding	
+Passive+Sg+Masc:jE	#;
+Passive+Sg+Fem:jE	#;
+Passive+Pl+Masc:jani	#;
+Passive+Pl+Fem:jani	#;
+Passive+Sg+Masc+ConPart:jI	#;
+Passive+Sg+Fem+ConPart:jI	#;
+Passive+Pl+Masc+ConPart:jI	#;
+Passive+Pl+Fem+ConPart:jI	#;
+Passive+Sg+Masc:bO	#;
+Passive+Sg+Fem:bI	#;
+Passive+Pl+Masc:bA	#;
+Passive+Pl+Fem:biyUN	#;

Figure III - 17: Infinitive and Passive Ending Sublexicon Definitions

Different morphological form of “likHu” generated by above defined morphology are shown in Table III-75.

Table III - 75: Different Morphological Forms of Transitive Verb “likHu”

Tag/Feature sequence	Intermediate Forms	Surface Form
likHu+Verb+Inf	likH a Nra	likHaNra
likHu+Verb+Imp+Caus1+Sg+2P	likH u	likHu
likHu+Verb+Imp+Caus1+Pl+2P	likH O	likHO
likHu+Verb+Imp+Polite+Caus1+Sg+2P	likH i ji	likHiji
likHu+Verb+Imp+Polite+Caus1+Pl+2P	likH i jO	likHijO
likHu+Verb+Imp+Des+Caus1+Sg+2P	likH i jANi	likHijANI
likHu+Verb+Imp+Des+Caus1+Pl+2P	likH i jaO	likHijaO
likHu+Verb+Imp+Caus2+2P	likH Ai	likHAI
likHu+Verb+Imp+Caus3+2P	likH ArAi	likHArAi
likHu+Verb+Imp+Caus4+2P	likH ArArAi	likHArArAi
likHu+Verb+Passive+Sg+Fem	likH i jE	likHijE
likHu+Verb+Passive+Sg+Masc	likH i jE	likHijE
likHu+Verb+Passive+Pl+Masc	likH i jani	likHijani
likHu+Verb+Passive+Pl+Fem	likH i jani	likHijani
likHu+Verb+Passive+Sg+Masc+ConPart	likH i jl	likHijl
likHu+Verb+Passive+Sg+Fem+ConPart	likH i jl	likHijl
likHu+Verb+Passive+Pl+Masc+ConPart	likH i jl	likHijl
likHu+Verb+Passive+Pl+Fem+ConPart	likH i jl	likHijl
likHu+Verb+Passive+Sg+Masc	likH i bO	likHibO
likHu+Verb+Passive+Sg+Fem	likH i bl	likHibl
likHu+Verb+Passive+Pl+Masc	likH i bA	likHibA
likHu+Verb+Passive+Pl+Fem	likH i biyUN	likHibiyUN
likHu+Verb+PresPart+Masc+Sg	likH andO	likHandO
likHu+Verb+PresPart+Masc+Pl	likH andA	likHandA
likHu+Verb+PresPart+Fem+Sg	likH andI	likHandI
likHu+Verb+PresPart+Fem+Pl	likH andyUN	likHandyUN
likHu+Verb+PastPart+Masc+Sg	likH yO	likHyO
likHu+Verb+PastPart+Masc+Pl	likH yA	likHyA
likHu+Verb+PastPart+Fem+Pl	likH yUN	likHyUN
likHu+Verb+PastPart	likH yala	Likhyala
likHu+Verb+PastPart+Fem+Sg	likH I	likHI
likHu+Verb+FutPart	likH aNrU	likHaNrU
likHu+Verb+FutPart+Sg	likH aNrU	likHaNrU
likHu+Verb+FutPart+Pl	likH aNrA	likHaNrA

Cont...

Table III-75: Continued.

likHu+Verb+VerbNoun	likH andaRa	likHandaRa
likHu+Verb+ConjPart	likH I	likHI
likHu+Verb+Aorist+Sg+1P	likH AN	likHAN
likHu+Verb+Aorist+Pl+1P	likH UN	likHUN
likHu+Verb+Aorist+Sg+2P	likH IN	likHIN
likHu+Verb+Aorist+Pl+2P	likH O	likHO
likHu+Verb+Aorist+Sg+3P	likH E	likHE
likHu+Verb+Aorist+Pl+3P	likH ani	Likhani
likHu+Verb+Future+Sg+Masc+1P	likH andasu	likHandasu
likHu+Verb+Future+Sg+Masc+2P	likH andIN	likHandIN
likHu+Verb+Future+Sg+Masc+3P	likH andO	likHandO
likHu+Verb+Future+Sg+Fem+1P	likH andiyasu	likHandiyasu
likHu+Verb+Future+Sg+Fem+2P	likH andiIN	likHandiIN
likHu+Verb+Future+Sg+Fem+3P	likH andI	likHandI
likHu+Verb+Future+Pl+Masc+1P	likH andAsIN	likHandAsIN
likHu+Verb+Future+Pl+Masc+2P	likH andaO	likHandaO
likHu+Verb+Future+Pl+Masc+3P	likH andA	likHandA
likHu+Verb+Future+Pl+Fem+1P	likH andyUNsIN	likHandyUN-sIN
likHu+Verb+Future+Pl+Fem+2P	likH andiaO	likHandiaO
likHu+Verb+Future+Pl+Fem+3P	likH andyUN	likHandyUN

4.2.3 Pronominal Suffixes

Definitions of sub lexicons which are used to define pronominal suffixes in Figure III-16 are shown in Figure III-18. The lexical entry “likHu+Verb:likH VerbStem1;” is further defined in VerbStem1 lexicon with following entries for pronominal suffixation:

+Psx:iy	PSuffix1;
+Psx+PastPart+Sg:iyo	PSuffix2;
+Psx+PastPart+Pl:iyA	PSuffix2;

First sub lexical entry is intermediate pronominal suffixation entry which is further inflected by PSuffix1 sub lexicon. For example, following sequence of tags with verb “likHu” will produce “likH-iy-ame” (I wrote).

likHu+Verb:liHh +Psx:iy +SSg+S1P+SMF+SObl+Sg+PastPart:ame #;

Here verb “likHu” will produce intermediate “likH” for which will be concatenated with intermediate suffix “-iy” which is further concatenated with ame to produce verb “likHyame” with pronominal suffixes. Various tags defined during morphological generation / analysis define the properties of pronoun which is hidden in verb form. For instance, “+SSg+S1P+SMF+SObl” tags define the subject pronoun properties which say that subject is singular, first person, oblique form. Different verb forms with pronominal suffixation generated are shown in Table III-76.

LEXICON PSuffix1	
+SSg+S1P+SMF+SObl+Sg+PastPart:ame	#;
+SSg+S2P+SMF+SObl+Sg+PastPart:aI	#;
+SSg+S3P+SMF+SObl+Sg+PastPart:AIN	#;
+SSg+S1P+SMF+SObl+Pl+PastPart:Ame	#;
+SSg+S2P+SMF+SObl+Pl+PastPart:aI	#;
+SSg+S3P+SMF+SObl+Pl+PastPart:AIN	#;
+SPl+S1P+SMF+SObl+Sg+PastPart:OsIN	#;
+SPl+S2P+SMF+SObl+Sg+PastPart:ava	#;
+SPl+S3P+SMF+SObl+Sg+PastPart:ane	#;
+SPl+S1P+SMF+SObl+Pl+PastPart:AsIN	#;
+SPl+S2P+SMF+SObl+Pl+PastPart:Ava	#;
+SPl+S3P+SMF+SObl+Pl+PastPart:Ane	#;
LEXICON PSuffix2	
+SSg+S1P+SObl+SMF:mAN	OBJPSuffix;
+SPl+S1P+SObl+SMF:sIN	OBJPSuffix;
LEXICON OBJPSuffix	
+OSg+O2P+OMF:e	#;
+OPl+O2P+OMF:va	#;
+OSg+O3P+OMF:se	#;
+OPl+O3P+OMF:ne	#;

Figure III - 18: Psuffix (Pronominal Suffix) Sublexicon Definitions

Table III - 76: Pronominal Suffix Surface Forms Generated by Pronominal Suffix Morphology

likHu+Verb+Psx + (feature tags)	Intermediate Forms	Surface Form
+PastPart+Sg+SSg+S1P+SObl +SMF+OSg+O2P+OMF	iyO mANe	likHiyOmANe
+PastPart+Sg+SSg+S1P+SObl +SMF+OPI+O2P+OMF	iyO mANva	likHiyOmANva
+PastPart+Sg+SSg+S1P+SObl +SMF+OSg+O3P+OMF	iyO mANse	likHiyOmANse
+PastPart+Sg+SSg+S1P+SObl +SMF+OPI+O3P+OMF	iyO mANne	likHiyOmANne
+PastPart+Sg+SPI+S1P+SObl +SMF+OSg+O2P+OMF	iyO sINe	likHiyOsINe
+PastPart+Sg+SPI+S1P+SObl +SMF+OPI+O2P+OMF	iyO sINva	likHiyOsINva
+PastPart+Sg+SPI+S1P+SObl +SMF+OSg+O3P+OMF	iyO sINse	likHiyOsINse
+PastPart+Sg+SPI+S1P+SObl +SMF+OPI+O3P+OMF	iyO sINne	likHiyOsINne
+PastPart+PI+SSg+S1P+SObl +SMF+OSg+O2P+OMF	iyA mANe	likHiyAmANe
+PastPart+PI+SSg+S1P+SObl +SMF+OPI+O2P+OMF	iyA mANva	likHiyAmANva
+PastPart+PI+SSg+S1P+SObl +SMF+OSg+O3P+OMF	iyA mANse	likHiyAmANse
+PastPart+PI+SSg+S1P+SObl +SMF+OPI+O3P+OMF	iyA mANne	likHiyAmANne
+PastPart+PI+SPI+S1P+SObl +SMF+OSg+O2P+OMF	iyA sINe	likHiyAsINe
+PastPart+PI+SPI+S1P+SObl +SMF+OPI+O2P+OMF	iyA sINva	likHiyAsINva
+PastPart+PI+SPI+S1P+SObl +SMF+OSg+O3P+OMF	iyA sINse	likHiyAsINse
+SMF+OPI+O3P+OMF	iyA sINne	likHiyAsINne
+SSg+S1P+SMF+SObl+Sg+PastPart	iy ame	likHiyame
+SSg+S2P+SMF+SObl+Sg+PastPart	iy al	likHiyal
+SSg+S2P+SMF+SObl+PI+PastPart	iy al	likHiyal
+SSg+S3P+SMF+SObl+Sg+PastPart	iy AIN	likHiyAIN
+SSg+S3P+SMF+SObl+PI+PastPart	iy AIN	likHiyAIN
+SSg+S1P+SMF+SObl+PI+PastPart	iy Ame	likHiyAme
+SPI+S1P+SMF+SObl+Sg+PastPart	iy OsIN	likHiyOsIN
+SPI+S2P+SMF+SObl+Sg+PastPart	iy ava	likHiyava
+SPI+S3P+SMF+SObl+Sg+PastPart	iy ane	likHiyanne
+SPI+S1P+SMF+SObl+PI+PastPart	iy AsIN	likHiyAsIN
+SPI+S2P+SMF+SObl+PI+PastPart	iy Ava	likHiyAva
+SPI+S3P+SMF+SObl+PI+PastPart	iy Ane	likHiyAne

Morphology of intransitive verb dORi (run) is defined by VerbStem2 sub lexicon shown in Figure VI-16. Table VI-11 and Table VI-12 show 65 different forms of verb dORi generated by FST.

```

LEXICON VerbStem2
0:a                                Infinitive;
0:i                                PassiveEnding;
+Imp+Sg+2P:i                         #;
+Imp+Pl+2P:O                         #;
+Imp+Polite+Sg+2P:iji                #;
+Imp+Polite+Pl+2P:ijo                #;
+Imp+Des+Sg+2P:ijANI               #;
+Imp+Des+Pl+2P:ijaO                #;
+PresPart+Masc+Sg:andO              #;
+PresPart+Masc+Pl:andA              #;
+PresPart+Fem+Sg:andI              #;
+PresPart+Fem+Pl:andyUN             #;
+PastPart+Masc+Sg:yO                #;
+PastPart+Masc+Pl:yA                #;
+PastPart:yO                          #;
+PastPart+Fem+Sg:Ia                #;
+PastPart+Fem+Pl:yUN                #;
+PastPart:yala                        #;
+FutPart+Sg:aNrU                     #;
+FutPart+Pl:aNrA                     #;
+VerbNoun:andaRa                     #;
+ConjPart:I                          #;
+Imp+Caus2+2P:Ai                     #; !may be caus1 or caus2 with null
+Imp+Caus3+2P:ArAi                  #;
+Imp+Caus4+2P:ArArAi                #;
+Aorist+Sg+1P:AN                     #;
+Aorist+Pl+1P:UN                     #;
+Aorist+Sg+2P:IN                     #;
+Aorist+Pl+2P:O                     #;
+Aorist+Sg+3P:E                     #;
+Aorist+Pl+3P:ani                   #;
+Future+Sg+Masc+1P:andasu           #;
+Future+Sg+Fem+1P:andiyasu          #;
+Future+Pl+Masc+1P:andAsIN          #;
+Future+Pl+Fem+1P:andyUNsIN         #;
+Future+Sg+Masc+2P:andIN            #;
+Future+Sg+Fem+2P:andiIN            #;
+Future+Pl+Masc+2P:andaO             #;
+Future+Pl+Fem+2P:andiaO            #;
+Future+Sg+Masc+3P:ando              #; !Present Participle form
+Future+Sg+Fem+3P:andi              #; !Present Participle form
+Future+Pl+Masc+3P:andA              #; !Present Participle form
+Future+Pl+Fem+3P:andyUN             #; !Present Participle form
+1P+PastInd+Sg:ayasu                #;
+1P+PastInd+Masc+Pl:ayAIN           #;
+1P+PastInd+Fem+Pl:ayooNsIN         #;
+2P+PastInd+Masc+Sg:ain              #;
+2P+PastInd+Fem+Sg:iIN              #;
+2P+PastInd+Masc+Pl:yA              #;
+2P+PastInd+Fem+Sg:yUN              #;
+3P+PastInd+Masc+Sg:yO              #;
+3P+PastInd+Fem+Sg:I                #;
+3P+PastInd+Masc+Pl:yA              #;
+3P+PastInd+Fem+Sg:yUN              #;

```

Figure III - 19: Sublexicon VerbStem2 (Intransitive Verbs) Definition

Table VI - 77: Different Morphological Forms of Intransitive Verb “dORi”

Tag/Feature sequence	Intermediate Forms	Surface Form
dORi+Verb+Inf	dOR a Nra	dORaNra
dORi+Verb+Passive+Sg+Fem	dOR i jE	dORijE
dORi+Verb+Passive+Sg+Masc	dOR i jE	dORijE
dORi+Verb+Passive+Pl+Masc	dOR i jani	dORijani
dORi+Verb+Passive+Pl+Fem	dOR i jani	dORijani
dORi+Verb+Passive+Sg+Masc+ConPart	dOR i jl	dORijl
dORi+Verb+Passive+Sg+Fem+ConPart	dOR i jl	dORijl
dORi+Verb+Passive+Pl+Masc+ConPart	dOR i jl	dORijl
dORi+Verb+Passive+Pl+Fem+ConPart	dOR i jl	dORijl
dORi+Verb+Passive+Sg+Masc	dOR i bO	dORibO
dORi+Verb+Passive+Sg+Fem	dOR i bl	dORibl
dORi+Verb+Passive+Pl+Masc	dOR i bA	dORibA
dORi+Verb+Passive+Pl+Fem	dOR i biyUN	dORibiyUN
dORi+Verb+Imp+Sg+2P	dOR i	dORi
dORi+Verb+Imp+Polite+Sg+2P	dOR iji	dORiji
dORi+Verb+Imp+Polite+Pl+2P	dOR ijO	dORijO
dORi+Verb+Imp+Des+Sg+2P	dOR ijANI	dORijANI
dORi+Verb+Imp+Des+Pl+2P	dOR ijaO	dORijaO
dORi+Verb+Imp+Pl+2P	dOR O	dORO
dORi+Verb+PresPart+Masc+Sg	dOR andO	dORandO
dORi+Verb+PresPart+Masc+Pl	dOR andA	dORandA
dORi+Verb+PresPart+Fem+Sg	dOR andI	dORandI
dORi+Verb+PresPart+Fem+Pl	dOR andyUN	dORandyUN
dORi+Verb+PastPart+Masc+Sg	dOR yO	dORYO
dORi+Verb+PastPart+Masc+Pl	dOR yA	dORYA
dORi+Verb+PastPart+Fem+Pl	dOR yUN	dORYUN
dORi+Verb+PastPart	dOR yala	dORYala
dORi+Verb+PastPart	dOR yO	dORYO
dORi+Verb+PastPart+Fem+Sg	dOR la	dORla
dORi+Verb+FutPart+Sg	dOR aNrU	dORaNrU
dORi+Verb+FutPart+Pl	dOR aNrA	dORaNrA
dORi+Verb+VerbNoun	dOR andaRa	dORandaRa
dORi+Verb+ConjPart	dOR I	dORI

Table VI - 78: Different Morphological Form of Intransitive Verb “dORi”

Tag/Feature sequence	Intermediate Forms	Surface Form
dORi+Verb+Imp+Caus2+2P	dOR Ai	dORAi
dORi+Verb+Imp+Caus3+2P	dOR ArAi	dORArAi
dORi+Verb+Imp+Caus4+2P	dOR ArArAi	dORArArAi
dORi+Verb+Aorist+Sg+1P	dOR AN	dORAN
dORi+Verb+Aorist+Pl+1P	dOR UN	dORUN
dORi+Verb+Aorist+Sg+2P	dOR IN	dORIN
dORi+Verb+Aorist+Pl+2P	dOR O	dORO
dORi+Verb+Aorist+Sg+3P	dOR E	dORE
dORi+Verb+Aorist+Pl+3P	dOR ani	dORani
dORi+Verb+Future+Sg+Masc+1P	dOR andasu	dORandasu
dORi+Verb+Future+Sg+Masc+2P	dOR andIN	dORandIN
dORi+Verb+Future+Sg+Masc+3P	dOR andO	dORandO
dORi+Verb+Future+Sg+Fem+1P	dOR andiyasu	dORandiyasu
dORi+Verb+Future+Sg+Fem+2P	dOR andiIN	dORandiIN
dORi+Verb+Future+Sg+Fem+3P	dOR andI	dORandI
dORi+Verb+Future+Pl+Masc+1P	dOR andAsIN	dORandAsIN
dORi+Verb+Future+Pl+Masc+2P	dOR andaO	dORandaO
dORi+Verb+Future+Pl+Masc+3P	dOR andA	dORandA
dORi+Verb+Future+Pl+Fem+1P	dOR andyUNsIN	dORandyUNsIN
dORi+Verb+Future+Pl+Fem+2P	dOR andiaO	dORandiaO
dORi+Verb+Future+Pl+Fem+3P	dOR andyUN	dORandyUN
dORi+Verb+1P+PastInd+Sg	dOR ayasu	dORayasu
dORi+Verb+1P+PastInd+Masc+Pl	dOR ayAIN	dORayAIN
dORi+Verb+1P+PastInd+Fem+Pl	dOR ayooNsIN	dORayooNsIN
dORi+Verb+2P+PastInd+Masc+Sg	dOR aIN	dORaIN
dORi+Verb+2P+PastInd+Fem+Sg	dOR iIN	dORiIN
dORi+Verb+2P+PastInd+Masc+Pl	dOR yA	dORYA
dORi+Verb+2P+PastInd+Fem+Sg	dOR yUN	dORYUN
dORi+Verb+3P+PastInd+Masc+Sg	dOR yO	dORYO
dORi+Verb+3P+PastInd+Masc+Pl	dOR yA	dORYA
dORi+Verb+3P+PastInd+Fem+Sg	dOR yUN	dORYUN
dORi+Verb+3P+PastInd+Fem+Sg	dOR I	dORI

4.3 Implementing Syntax in XLE

Various syntactic constructions of Sindhi grammar are implemented in XLE by defining Sindhi LFG rules. Morphology defined in compiled finite state transducers of lexicon developed in LEXC is used by these rules to parse Sindhi sentences. This section discusses the Sindhi LFG syntax implementation details which include nominal and verbal syntax. Following sub sections discuss the XLE grammar file structure, morphology syntax interface and implementation details of nominal and verbal elements of Sindhi LFG grammar.

4.4 XLE Grammar File Structure

Figure III-20 shows the grammar file structure of Sindhi LFG grammar file. There are four different sections in this grammar file. Sections start with a section header and ends with four dashes (----). Four elements in every section header include version, followed by language, component and XLE version number. First section can be explained as version RAHMAN, Language SINDHI, component CONFIG and XLE version (1.0). Remaining sections are TEMPLATES, RULES and LEXICON respectively. Config section contains configuration details of the grammar including the root category of syntax rules, (S for sentence here), external files included in the grammar (no file in this case), lexicon, rules and templates specifications and other configuration details. Detailed specifications of configuration section can be found at (Crouch, D., et al., 2011); however, individual configuration elements will be discussed in detail wherever necessary in subsequent sections.

Next section is morphology section which defines the morphology interfacing and tokenization files. List of finite state tokenizers for parsing and generation (default tokenizers are used here) and finite state transducer files which define the inflectional morphology of Sindhi (discussed in Section 2) are

```

RAHMAN SINDHI CONFIG (1.0)
ROOTCAT S.
FILES .
LEXENTRIES (RAHMAN SINDHI).
RULES (RAHMAN SINDHI).
TEMPLATES (RAHMAN SINDHI).
GOVERNABLERELATIONS SUBJ OBJ OBJ2 OBL OBL-?+ COMP XCOMP.
SEMANTICFUNCTIONS ADJUNCT TOPIC.
NONDISTRIBUTIVES NUM PERS.
EPSILON e.
OPTIMALITYORDER NOGOOD.
-----
RAHMAN SINDHI MORPHOLOGY (1.0)
P!default-parse-tokenizer.fsmfile G!default-gen-tokenizer.fst

ANALYZE:
SindhiAdjectivesV1.fst
SindhiNounsV2.fst
SindhiPronounsV1.fst
SindhiVerbV4.fst
SindhiAdverbs.fst
-----
RAHMAN SINDHI TEMPLATES (1.0)
-----
RAHMAN SINDHI RULES (1.0)
-----
RAHMAN SINDHI_MORPH RULES (1.0)
-----
RAHMAN SINDHI LEXICON (1.0)
-----
```

Figure III - 20: Sindhi LFG Grammar File Structure

placed here. Templates section follows the morphology section. Templates used in grammar are defined in templates section. Rules section holds the LFG grammar rules. Sindhi morph rules section will define the morphology

interacting rules to extract various attributes defined by finite state morphology transducers. In lexicon section, full form lexicon and morphology based lexicon is defined by using the attributes extracted from finite state transducers.

4.5 Implementing Nominal Syntax

As discussed in Chapter V nominal syntax include nouns, pronouns, adjectives, adverbs and phrases constituted by these elements. NP constructions and their analysis in XLE is discussed in detail in following sections.

4.5.1 Nouns in Noun Phrase

Consider following definition of NP in LFG.

$NP \rightarrow N : ^= ! .$

A noun phrase can be defined as noun only; and when a noun CHOkirO is analyzed or parsed by using above rule it produces a parse tree (see Figure III-21 a) and F-structure (see Figure III-21 b). F-structure represents various attributes of noun along-with their values generated by morphological analyzer. F-structure analysis shows that “CHOkirO” is animate, common count noun, in nominative case, with masculine gender and singular number. Syntax analysis of two more morphological forms of “CHOkirO” is shown in Figure III-22. Here (a) shows the analysis of oblique form of “CHOkirO” i.e. “CHOkirE”, and (b) shows two different possible analyzes of “CHOkirA” where “a:1” is vocative case in singular form and “a:2” is nominative case with plural number.

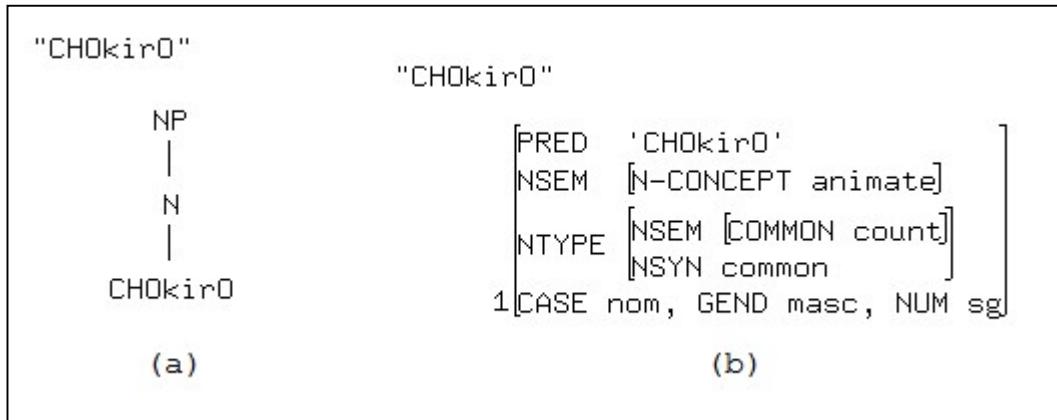


Figure III - 21: Parse Tree and F-structure of Noun "CHOkirO"

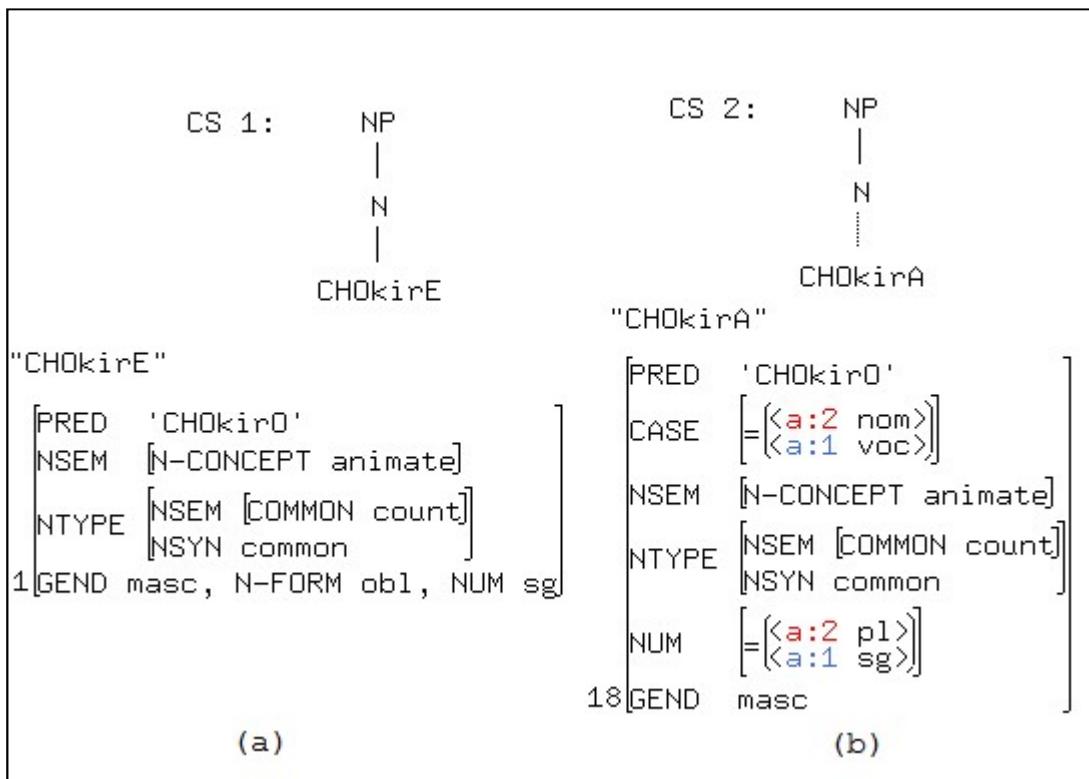


Figure III - 22: F-structure Analyses of Different Forms of Common Noun "CHOkirO"

4.5.2 Pronouns in Noun Phrase

Pronouns can act as standalone elements of NP or with other NP constituents. Following phrase structure rules define different pronoun types as NPs.

$NP \rightarrow N \mid P\text{Pron} \mid NP\text{Pron}(N)$

$NP\text{Pron} \rightarrow D\text{Pron} \mid Wh\text{Pron} \mid Ind\text{Pron} \mid Ref\text{Pron}$.

Figure III-23 shows the equivalent LFG definition of above phrase structure rules. Personal pronouns can appear standalone in NPs; however, other pronoun types specified in LFG rule can appear optionally with noun. Pronouns when appear with nouns in an NP they act as determiners; these determiners are placed within SPEC (specifier) f-structure. Figure III-24 shows LFG analyses of different personal pronouns; (a) shows personal pronoun “hl” (he) with masculine gender, singular number, 3rd person and nominative case. In the same way ‘ihA’ (it) and “uhE” (they) are shown in (b) and (c) respectively. Figure III-25 shows the NP construction with a demonstrative pronoun “hUa” (that) and noun “CHOKirla” (girl). F-structure analysis shows that “hUa” is determiner with different demonstrative pronoun attributes. NP constructions with reflexive and indefinite pronouns are shown in Figure III-26.

```

NP --> { N:^=! | PersPRON: ^=! | "Standalone Noun or Personal
Pronoun"
PRON:(^ SPEC DET)=!
{ (! PRON-TYPE)= demon |
(! PRON-TYPE)=whp |
(! PRON-TYPE)=indefinite |
(! PRON-TYPE)=reflexive };
(N:^=!) }. "Specific pronoun Followed by optional noun"

```

Figure III - 23: Noun Phrase Definition with Pronominal Elements in LFG

"hI"	"ihA"	"uhE"
NP	NP	NP
⋮	⋮	⋮
PersPRON	PersPRON	PersPRON
PRON	PRON	PRON
hI	ihA	uhE
(a)	(b)	(c)
"hI"		
[PRED 'hI' NTYPE [NSYN pronoun] 1[CASE nom, GEND masc, NUM sg, PERS 3, PRON-TYPE personal]]		
(a)		
"ihA"		
[PRED 'ihA' NTYPE [NSYN pronoun] 1[CASE nom, GEND fem, NUM sg, PERS 3, PRON-TYPE personal]]		
(b)		
"uhE"		
[PRED 'uhE' NTYPE [NSYN pronoun] 1[CASE nom, GEND fem, NUM pl, PERS 3, PRON-TYPE personal]]		
(c)		

Figure III - 24: LFG Analysis of Different Personal Pronouns

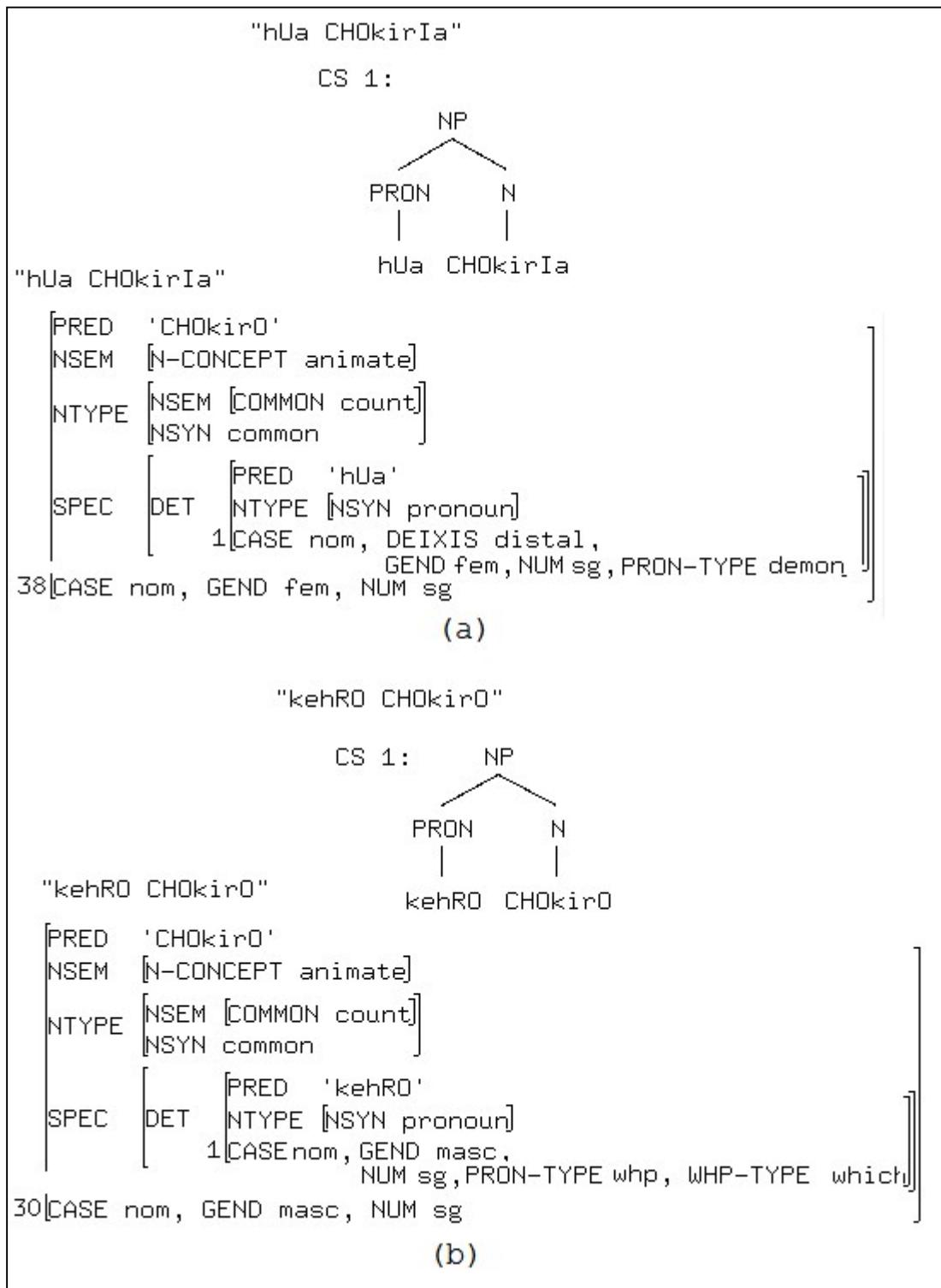


Figure III - 25: NP construction with demonstrative and Wh-pronouns

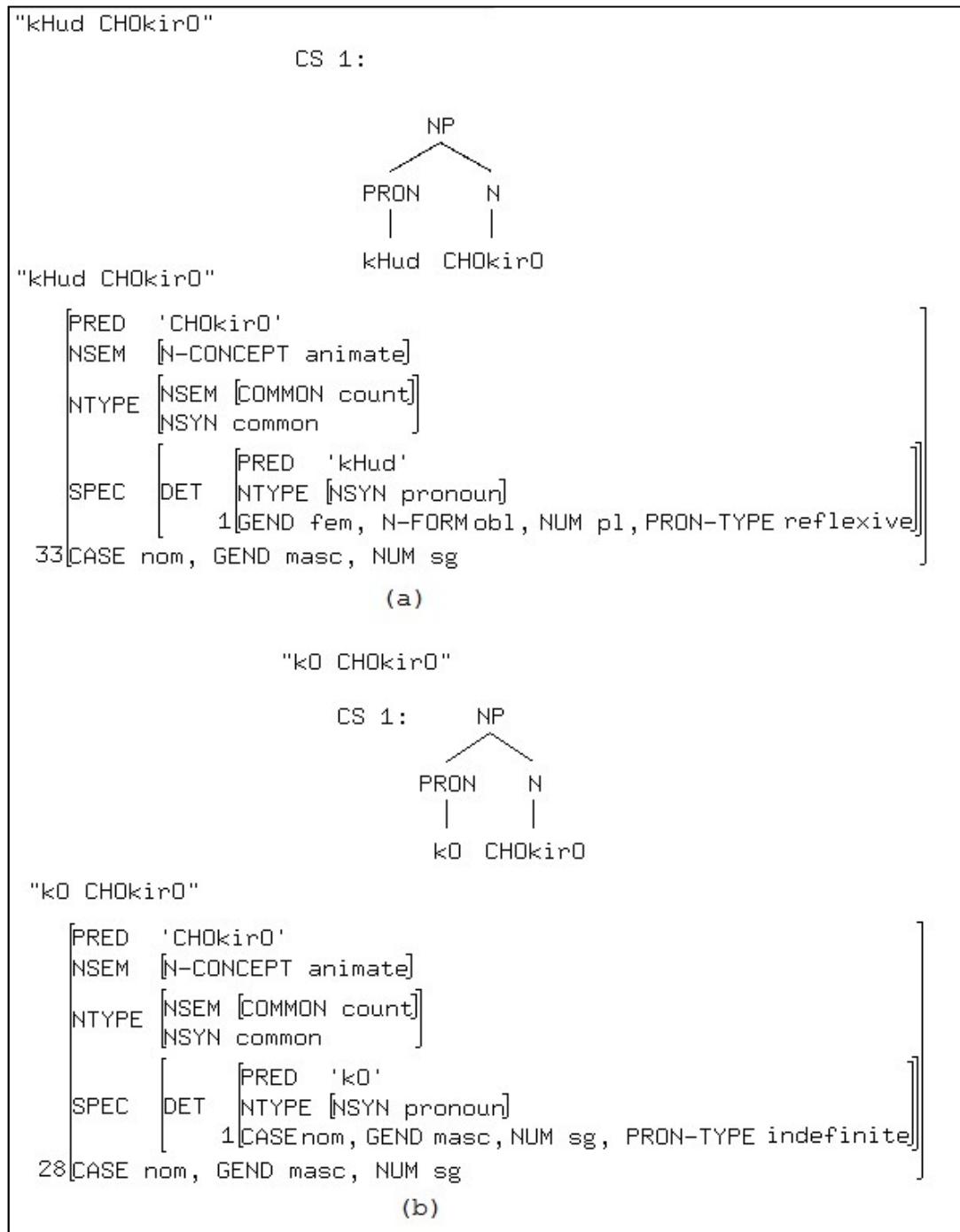


Figure III - 26: NP construction with Reflexive and Indefinite Pronouns

4.5.3 Adjectives in Noun Phrase

As discussed earlier there can be any number of adjective phrases before head noun in a noun phrase. An adjective phrase can be defined as single adjective optionally followed by any number of adjectives followed by a conjunction and an adjective. LFG c-structure rules for adjective phrase (ADJP) are shown below:

```

ADJP --> ADJ: ! $ ^
  (^ NUM)=( ! NUM)
  (^ GEND)=( ! GEND)
  (^ CASE)=( ! CASE);
(ADJ*: ! $ ^
  (^ NUM)=( ! NUM)
  (^ GEND)=( ! GEND)
  (^ CASE)=( ! CASE);
CONJ
ADJ: ! $ ^
  (^ NUM)=( ! NUM)
  (^ GEND)=( ! GEND)
  (^ CASE)=( ! CASE) .

```

Adjectives belong to adjunct sets and these sets are formed with “! \$ ^” notation which is equivalent of “ $\downarrow \in \uparrow$ ” in LFG theory. Annotations after “! \$ ^” in every ADJ occurrence will ensure number, gender and case agreement in coordinated adjective phrase. Sample analysis of single adjective ADJP, and coordinated adjectives ADJP is shown in Figure III-27 and Figure III-28 respectively. NP which contains ADJP can be defined in LFG rules as given below:

```

NP --> ADJP*: !$ (^ ADJUNCT)
  (^ NUM)=( ! NUM)
  (^ GEND)=( ! GEND)
  (^ CASE)=( ! CASE);
N:^=! .

```

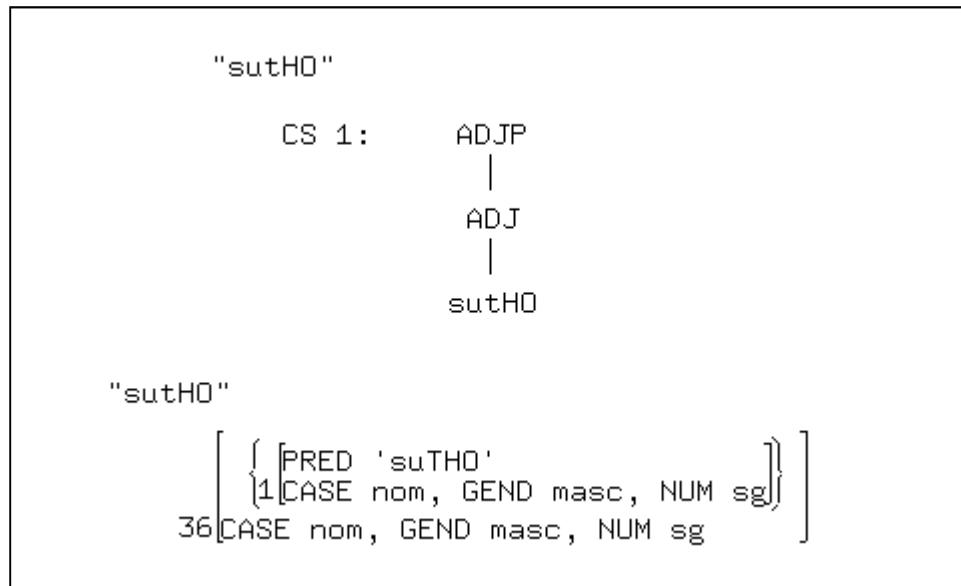


Figure III - 27: Simple ADJP (Adjective Phrase) Analysis

There can be any number of adjective phrases in a noun phrase and every adjective phrase belongs to set of adjuncts of NP. Analysis of sample noun phrase containing a coordinated adjective phrase with three adjectives is shown in Figure III-29.

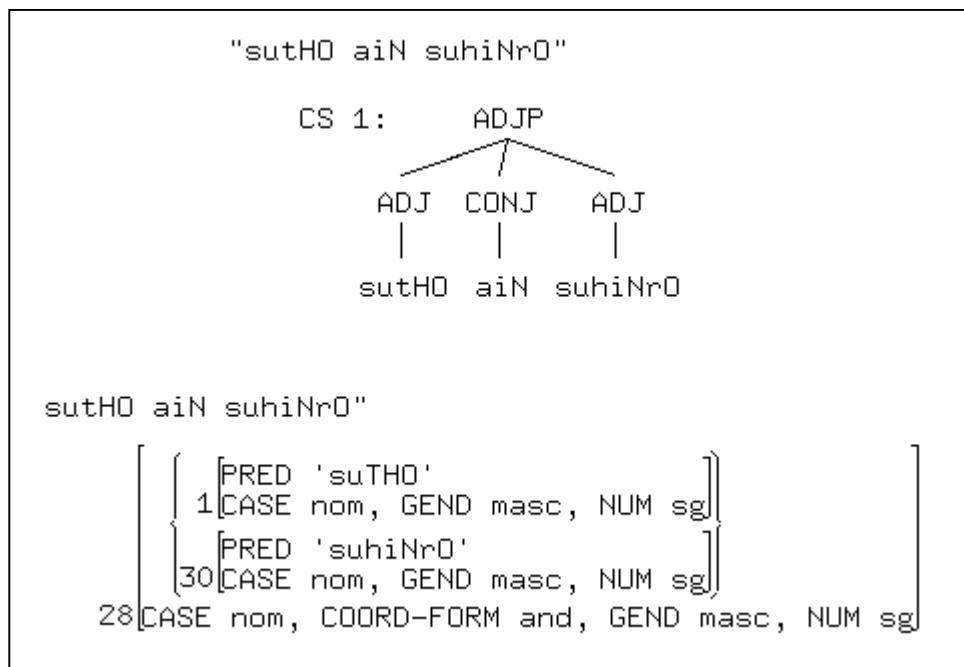


Figure III - 28: Coordinated ADJP Analysis.

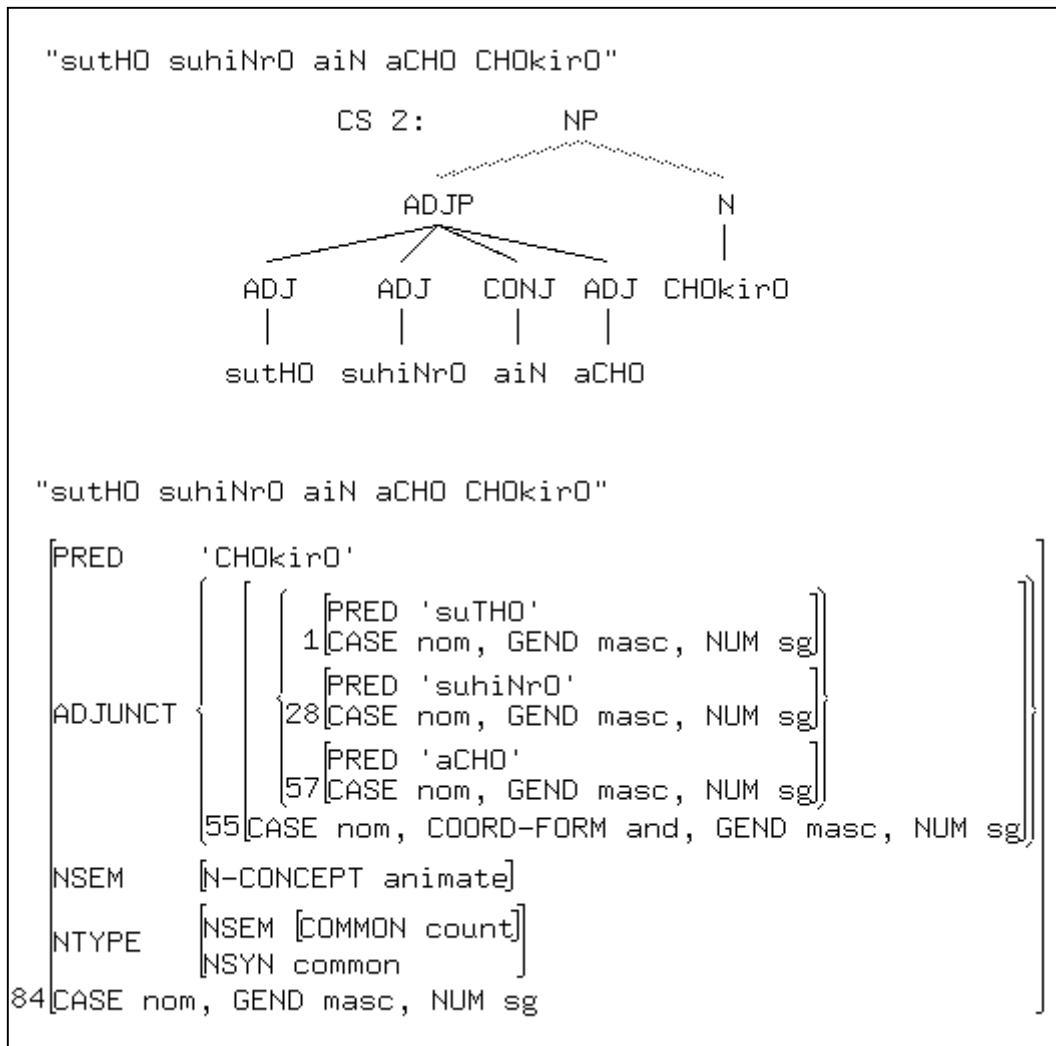


Figure III - 29: NP Analysis with Three Coordinated Adjectives.

4.5.4 Noun Phrases with Postpositional Phrases

NPs can also be modified by postpositional phrases these postpositional phrases contain noun phrases themselves. This indirect recursion of NP produces complex NPs containing complex postpositional phrases. Consider following LFG rules for postpositional phrase PPP.

```

PPP --> NP: !$^ {
    (! N-FORM)=obl (! CASE) ~= nom |
    (! NTYPE NSYN) =proper};
    PP.
  
```

Postpositional phrases also belong to set of adjuncts; and they are formed either by proper nouns or common nouns in oblique form. The annotation (!CASE) ~ nom here just ensures that proper nouns are not assigned oblique noun form in nominal case. Figure III-30 shows a sample noun phrase with postpositional phrase.

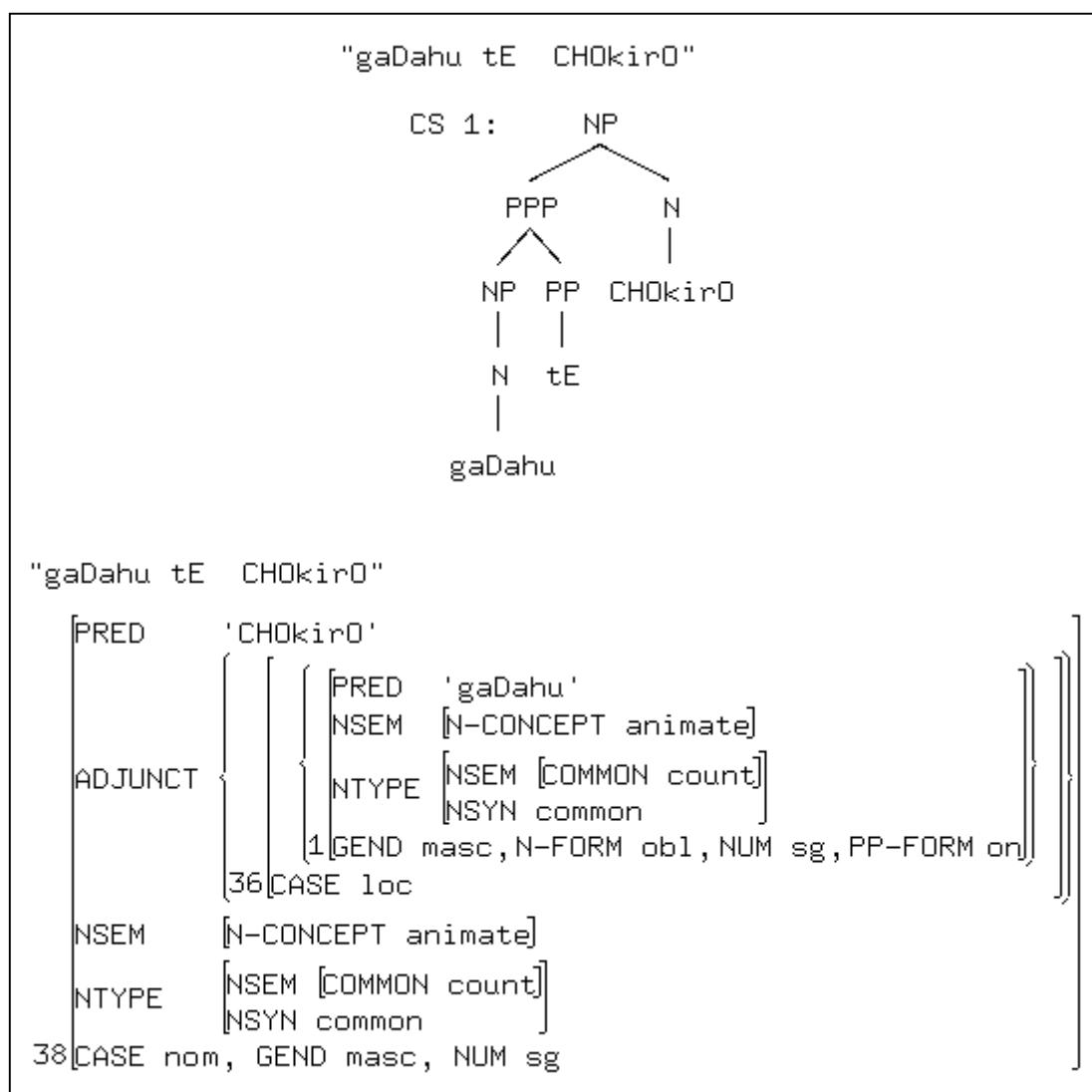


Figure III - 30: An NP with Postpositional Phrase.

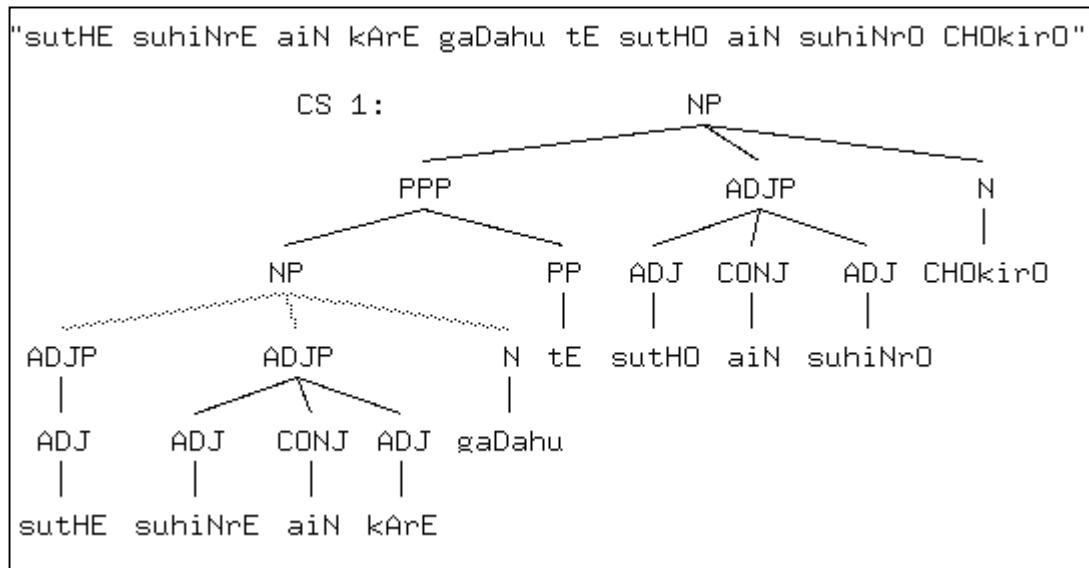


Figure III - 31: Parse Tree of Complex NP Construction Using Postpositional Phrase having Embedded NP.

As discussed above, PPP is defined in terms of NP; more complex NP constructions can also take place. Figure III-31 shows parse tree of more complex NP with adjective phrase which contains an NP with adjective phrase within postpositional phrase. It may also be noted that postpositions following nouns mark different cases of nouns. Postposition "tE" is used as locative case marker. This can be seen in f-structure of Figure III-32. This syntactical case marking of nouns by using postpositional case markers is discussed in following section.

4.5.5 Case Marking of Nouns / Noun Phrases

Apart from default nominative case, vocative case is also marked morphologically. However, all other cases discussed in Section 3.2 are marked syntactically by using case markers. F-structure in Figure III-33 shows nominative and vocative case analysis of common noun "CHOkirA". These

cases are morphological cases therefore no case marker is used here "CHOkirA" can either be in vocative case with singular number (a:1) or nominative case with plural number (a:2).

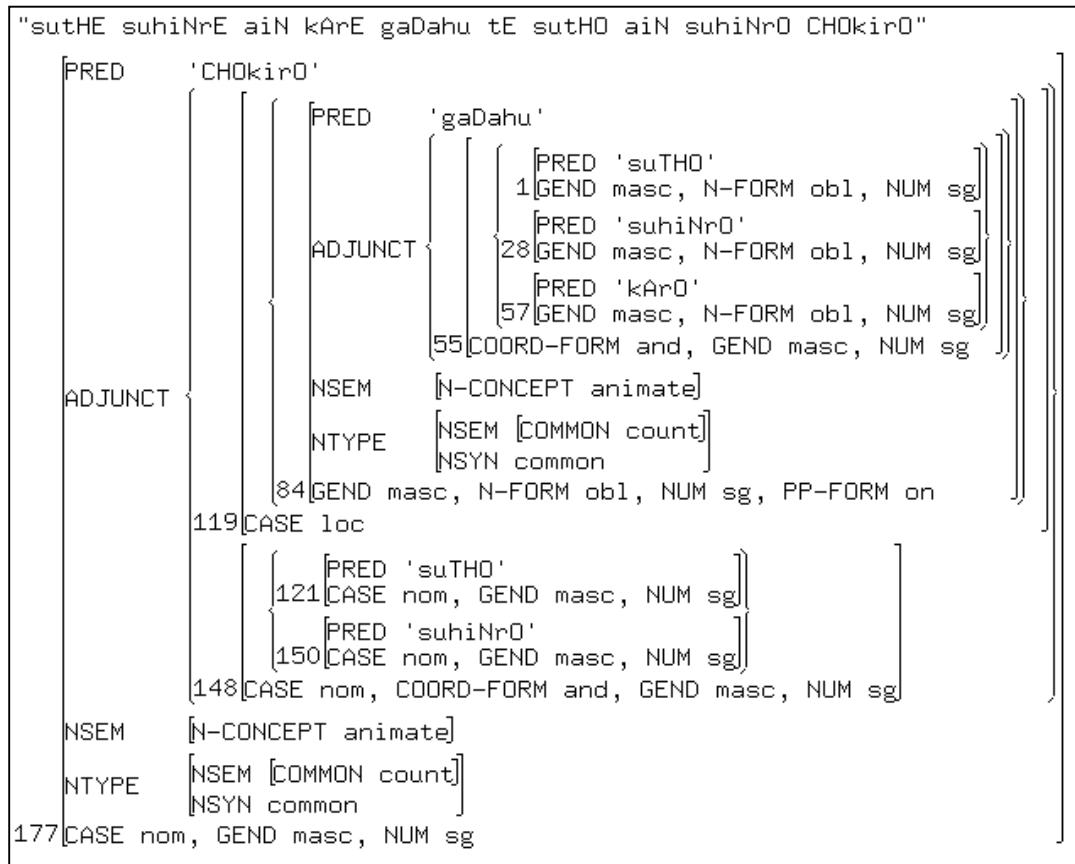


Figure III - 32: F-structure Analysis of Complex NP Construction Using Postpositional Phrase having Embedded NP.

F-structure chart "CHOkirA"	
"CHOkirA"	PRED 'CHOkirO'
CS 1:	CASE [$\begin{cases} \langle a:2 \text{ nom} \rangle \\ \langle a:1 \text{ voc} \rangle \end{cases}$]
NP	NSEM [N-CONCEPT animate]
	NTYPE [NSEM [COMMON count] NSYN common]
N	NUM [$\begin{cases} \langle a:2 \text{ pl} \rangle \\ \langle a:1 \text{ sg} \rangle \end{cases}$]
⋮	
CHOkirA	18[GEND masc]

Figure III - 33: Case Analysis of Common Noun Form "CHOkirA"

4.5.6 Case Phrase (KP)

Syntactic case marking is done via case phrase (KP). In KP noun phrase is marked with syntactic case. LFG definition of KP is given below:

```
KP --> NP: { (! N-FORM)=c obl |
    (! NTYPE NSYN)= proper} ^=! ;
K: ^=! .
```

Here KP is defined as an NP followed by case marker K. Case marker K is a postposition used as case marker and is defined in LFG lexicon section. For syntactic case marking noun phrase can either be in oblique morphological form or it is a proper noun. Different NP case marking examples via KP are discussed below.

Figure III-34 shows c-structure and f-structure analysis of accusative and dative case marking of common noun "CHOkirO". "khE" is used as accusative as well as dative case marker. Case is finalized at verbal subcategorization level; "CHOkirO" is said to be in accusative case if it is direct

object and in dative case if its indirect object. However, as this is standalone KP therefore case attribute has both values “acc” and “dat”. LFG lexicon entry of case marker “khE” is shown below:

```
khE      K * (^ PP-FORM)=khe
          { (^ CASE)=acc | (^ CASE)= dat} .
```

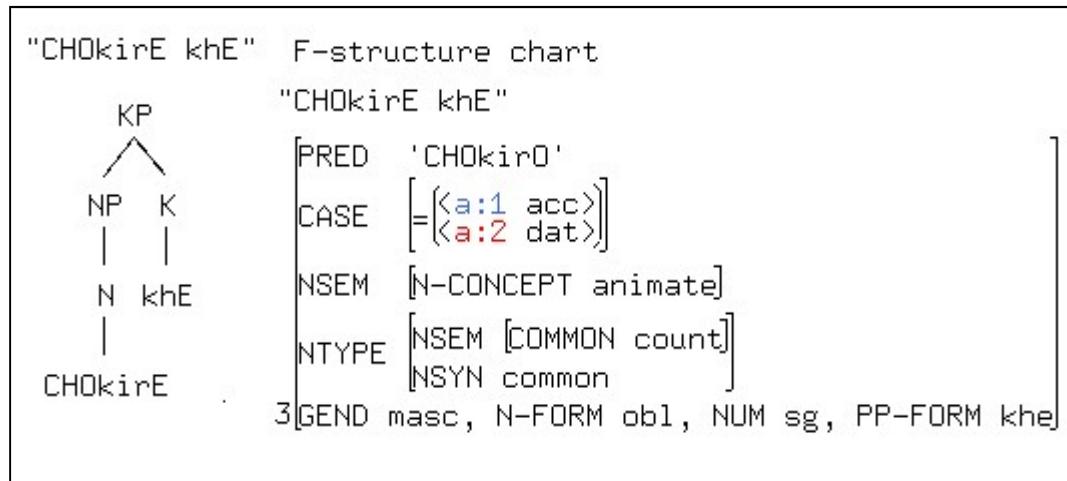


Figure III - 34: Accusative and Dative Case Marking of Common Noun
“CHOkirO”

Participant and instrumental cases are formed by “sAN” marker. In case if NP is an animate noun or pronoun “sAN” will act as participant case marker and if NP is inanimate noun then “sAN” will be instrumental case marker. LFG lexicon entry of “sAN” as case marker K which places constraints on noun type and concept and assigns case is given below:

```
sAN K * (^ PP-FORM)=with
          { { (^ NSEM N-CONCEPT)=c animate |
              (^ NTYPENSYN)=c pronoun}
            (^ CASE)=part |
            (^ NSEM N-CONCEPT)=c inanimate
            (^ CASE)=inst} .
```

Figure III-35 shows examples of participant and instrumental case with pronoun, animate noun and inanimate noun.

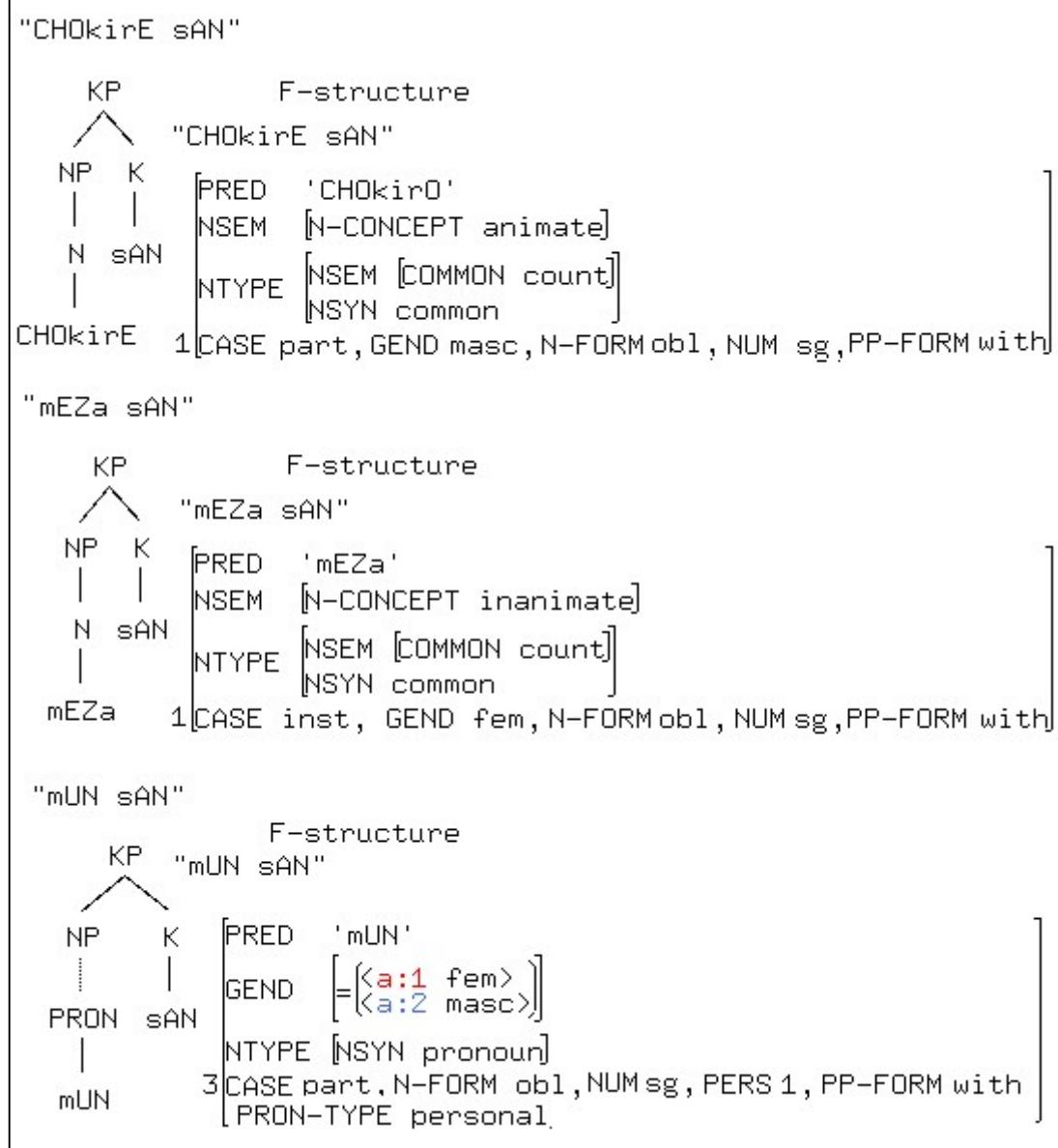


Figure III - 35: Participant and Instrument Case Marking with Case Marker "sAN".

“tE” and “mEN” are two locative case markers used in Sindhi. LFG lexicon entries of “tE” and “mEN” are shown below. Figure III-36 shows locative case marking examples marked by “tE” and “mEN”.

tE K * (^ PP-FORM) = on
(^ CASE)=loc.

mEN K * (^ PP-FORM) = in
(^ CASE)=loc.

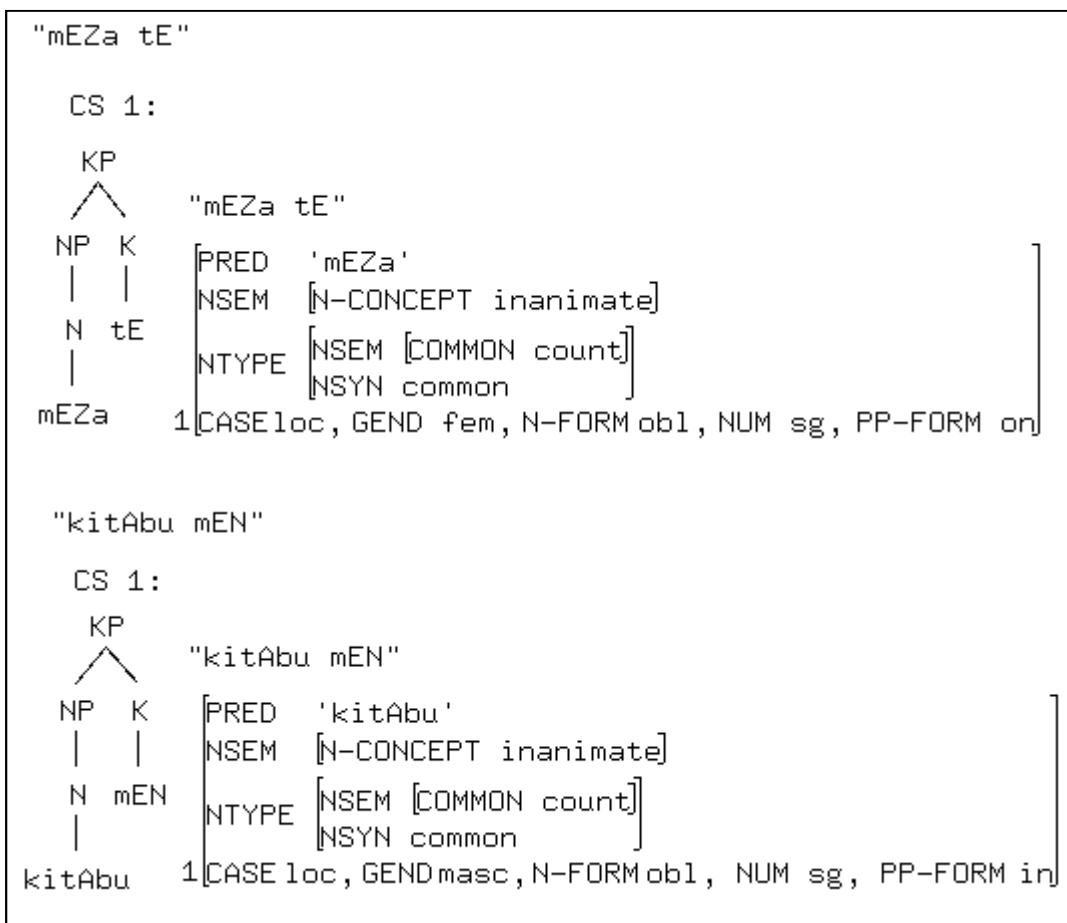


Figure III - 36: Locative Case Marking with “tE” and “mEN” Markers.

Ablative and agentive cases are marked by case marker “khAN”. Sometimes ablative case is marked morphologically. When “khAN” is used as case marker the LFG lexicon entry marks ablative as well as agentive case

this ambiguity is resolved at verbal subcategorization level when functional category of noun is decided. Figure III-37 (a) shows the analysis of morphological ablative case of common noun "gHaru". In Figure III-37 (b) f-structure analysis of agentive and ablative case marking with "khAN" marker is shown. Lexical entry of khAN case marker is given below.

```
khAN K * (^ PP-FORM)=khan
      { (^ CASE)=abl | (^ CASE)=agent} .
```

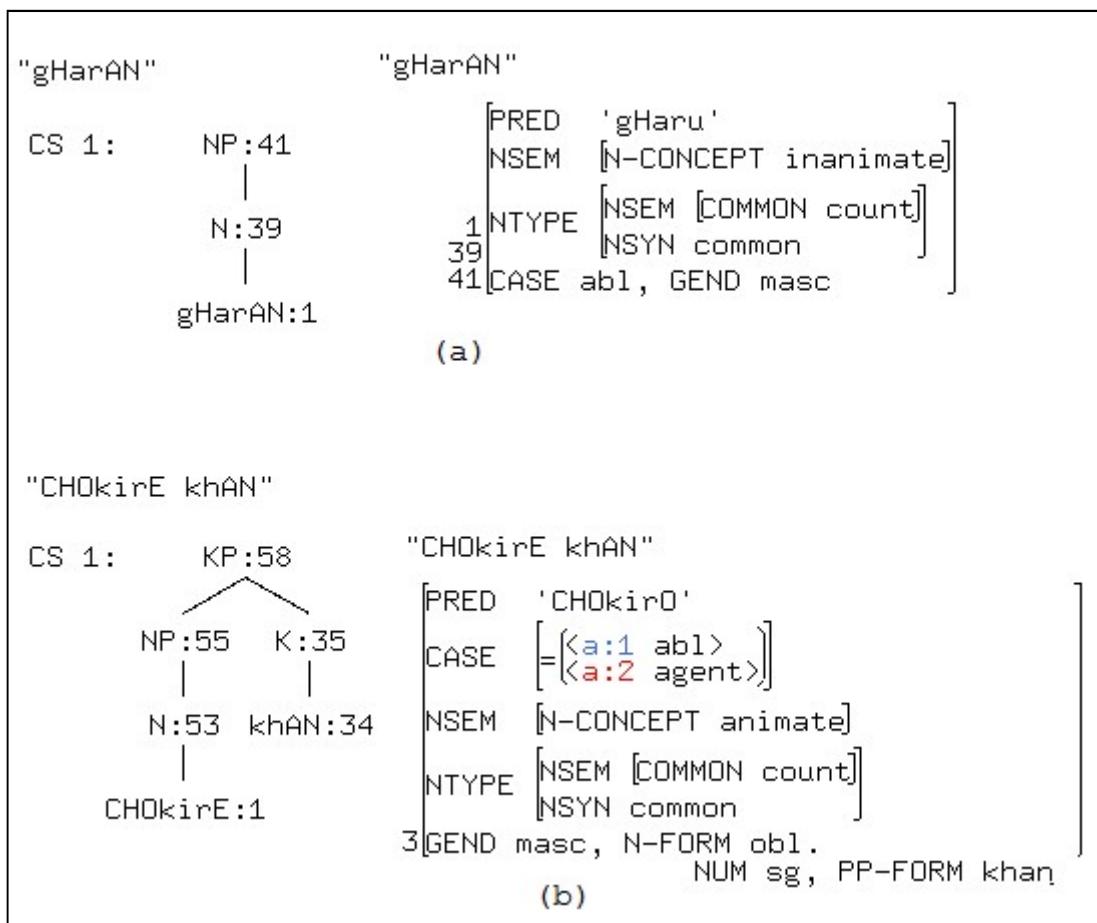


Figure III - 37: Ablative and Agentive Case Marking.

4.5.7 Possessive / Genitive Case

As discussed in section 3.2.11, possessive case marking is bit different than other syntactic cases in Sindhi. Possessive case marker inflects and must agree in number and gender (sometimes form must also agree) with possessed noun phrase. A separate possessive case phrase is therefore defined. LFG definition of possessive case phrase KPPoss is given below.

```
KPPoss --> NP: { (! N-FORM)=c obl |
                    (! NTYPE NSYN)= proper} ^=! ;
KPoss: ^=!. 
```

LFG lexicon entries of different inflections of KPoss (possessive case marker) “jO” are given below. It may be noted that extra attributes K-NUM, K-GEND, and K-FORM are introduced here to reflect the possessive case marker attributes to be agreed with possessed noun in noun phrase.

```
jO      KPoss * (^ PP-FORM)=of
          (^ K-NUM)=sg
          (^ K-GEND)=masc
          (^ K-FORM)=nom
          (^ CASE)=gen. 
```

```
jA      KPoss * (^ PP-FORM)=of
          { (^ K-NUM)=pl (^ K-FORM)=nom |
            (^ K-NUM)=sg (^ K-FORM)=voc}
          (^ K-GEND)=masc
          (^ CASE)=gen. 
```

```
jUN     KPoss * (^ PP-FORM)=of
          (^ K-NUM)=pl
          (^ K-GEND)=fem
          (^ K-FORM)=nom
          (^ CASE)=gen. 
```

```
jI      KPoss * (^ PP-FORM)=of
          (^ K-NUM)=sg
          (^ K-GEND)=fem
          { (^ K-FORM)=nom | (^ K-FORM)=voc}
          (^ CASE)=gen. 
```

```
jE      KPoss * (^ PP-FORM)=of
        (^ K-NUM)=sg
        (^ K-GEND)=masc
        (^ K-FORM)=obl
        (^ CASE)=gen.
```

First and second person pronouns also have morphological possessive case.

Figure III-38 shows the f-structure analysis of morphological genitive case of first and second person pronouns “mUN” and “tUN”.

"mUhInjO"	<pre> 1[PRED 'mUN' 52[NTYPE [NSYN pronoun] 54[CASE gen, GEND masc, K-GEND masc, K-NUM sg, NUM sg, PERS 1, PRON-TYPE personal]</pre>
"tuhiNjE"	<pre> 1[PRED 'tUN' 55[NTYPE [NSYN pronoun] 57[CASE gen, GEND masc, K-FORM abl, K-GEND masc, K-NUM sg, NUM sg, PERS 2, PRON-TYPE personal]</pre>

Figure III - 38: Morphological Genitive Case Marking of First and Second Person Pronouns.

Figure III-39 shows the parse tree and f-structure analysis of possessive case phrase “CHOkirani jA”. Here plural noun is marked with genitive plural masculine case. Noun phrase definition is now extended and contains possessive case phrase as well. Following LFG rule (after ellipses) is just fragment of NP definition which defines possibility of NP in possessive case phrase terms.

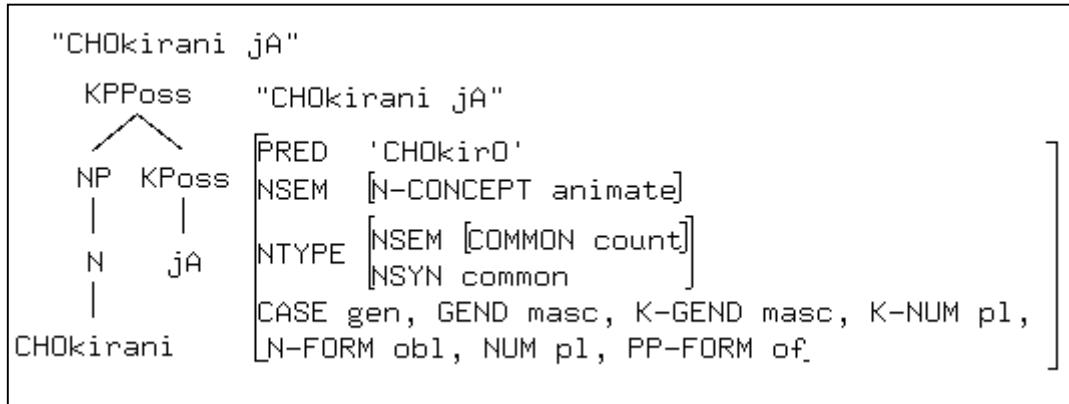


Figure III - 39: Plural Genitive Case Marking with Marker “jA”.

```

NP --> { ... | KPPoss: (^ SPEC)=! ;
NP: (! NUM)=(^ SPEC K-NUM)
      (! GEND)=(^ SPEC K-GEND)
      { (^ SPEC K-FORM)=c obl
        (! N-FORM)=c obl |
        (^ SPEC K-FORM)=c nom
        (! CASE)=c nom |
        (^ SPEC K-FORM)=c voc
        (! CASE)=c voc}
      ^=!
} . "Noun Phrase Definition"

```

Above rule defines noun phrase as possessive phrase specifier (SPEC) followed by possessed NP. As discussed above possessed NP must agree in number, gender and case / noun-form with case marker. Case marker attributes are reflected in K- prefixed attributes by KPPoss and are part of SPEC. Annotations following possessed NP are ensuring this agreement. Figure III-40 shows f-structure analysis of sample NP containing possessive specifier. Figure III-41 shows f-structure analysis which shows inconsistency problem in agreement as case marker “jl” is feminine and “kitAbu” is masculine. More complex NP constructions are now possible and nested NPs with nested possessive case phrases can be generated. Figure III-42 shows a parse tree

of such nesting with three levels. Figure III-43 shows the f-structure analysis of same NP.

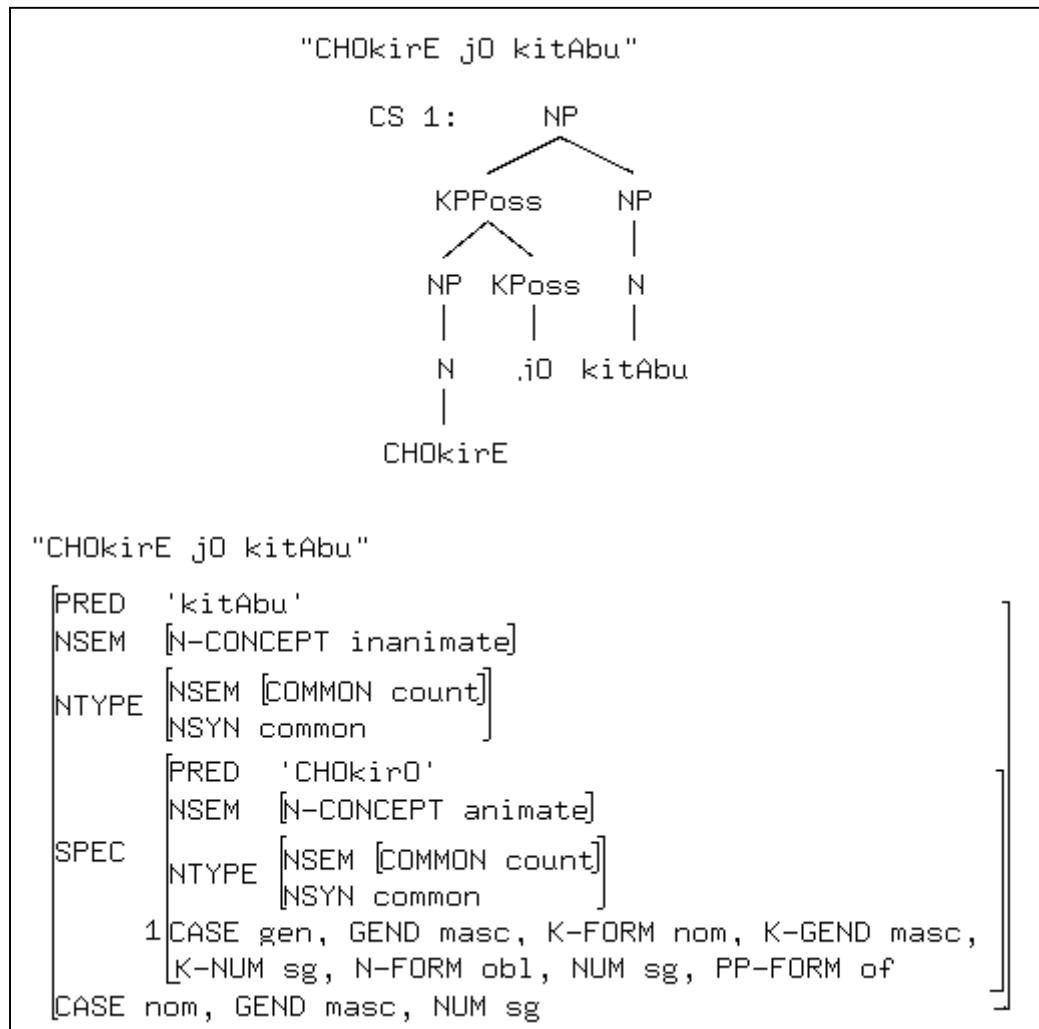


Figure III - 40: NP Example with Possessive Specifier.

```

F-structure #1 (INCONSISTENT)
"CHOkirE jI kitAbu"

PRED   'kitAbu'
NSEM   [N-CONCEPT inanimate]
NTYPE  [NSEM [COMMON count]
        NSYN common
        PRED   'CHOkirO'
        NSEM   [N-CONCEPT animate]
        NTTYPE [NSEM [COMMON count]
                  NSYN common
                  K-GEND [=fem]
                           =masc]
        1[CASE gen, GEND masc, K-NUM sg,  PP-FORM of]
        GEND  [=fem]      N-FORM obl, NUM sg,
36N-FORM obl, NUM sg

```

Figure III - 41: Inconsistent F-structure Due to Different Genders of Maker and Possessed Noun.

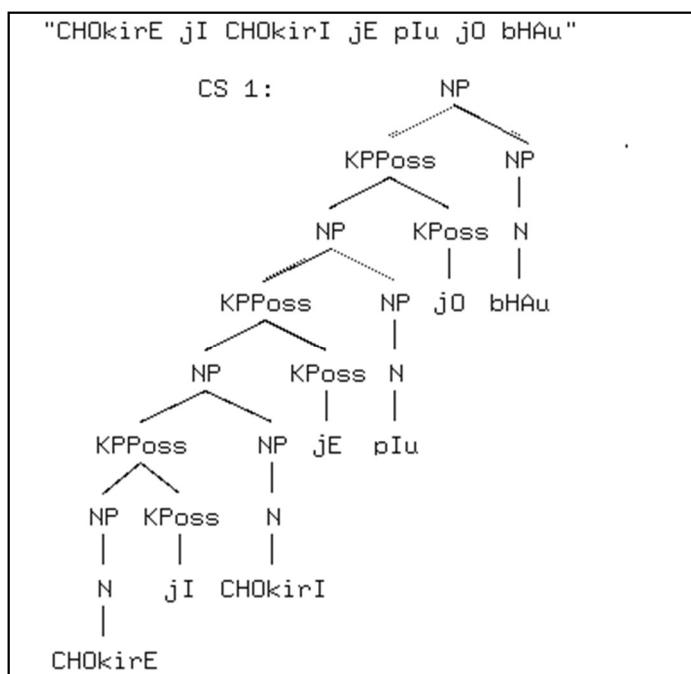


Figure III - 42: Nested NP Example.

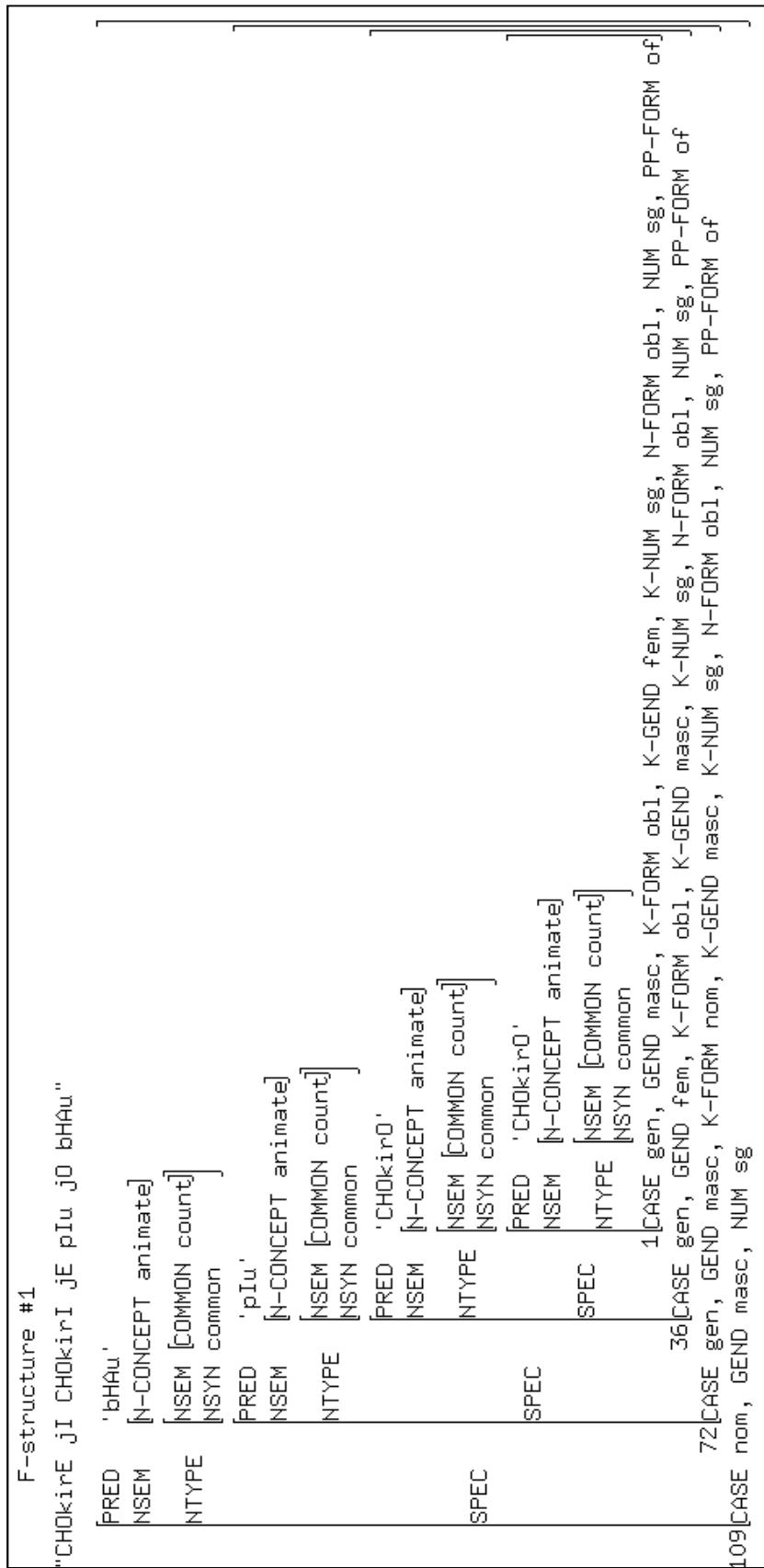


Figure III - 43: F-structure Analysis of Nested NP.

4.6 Implementing Verbal Syntax

Sentence can have nominal and verbal elements. Previous section discussed nominal syntax in detail. Verbal elements include verbs which subcategorize (require arguments) for different grammatical functions which include: subject (SUBJ), object (OBJ), secondary object (OBJ2), oblique (OBL), complement (COMP) and open complement XCOMP. Noun phrases (including all nominal elements) either define these functions or play essential role in their definition within a sentence. Sentence constituents therefore include verbs, their arguments and adjunct (ADJUNCT) elements which do not subcategorize for verbs. Verbs can either be predicative verbs (main verbs and copula verbs), modal verbs or auxiliary verbs. In Sindhi, main, auxiliary, and modal verbs are combined to make verbal complex. Auxiliaries are also used to mark tense, aspect and mood. In subsequent sections LFG implementation details of verbal subcategorization for different grammatical functions, verbal complex, tense-aspect-mood marking and other sentence constituents are discussed in detail.

4.6.1 Subject

As discussed in section 3.3 verbs subcategorize for subjects. However, this is not always true as some passive forms of verbs don't require subjects and subjects are replaced with NULL arguments. Following subcategorization frame is an example of intransitive verb dORi (run) which requires subject argument. The lexical entry calls V-SUBJ template to define its subcategorization frame by passing the stem as parameter of template. V-

SUBJ further calls PASS (passive) template which defines the passive construction where subject becomes NULL.

```
dORi    V XLE @ (V-SUBJ %stem) .
```

```
V-SUBJ(_P_)= { @ (PASS (^ PRED)=' _P_<(^ SUBJ)>' )
| (^ PRED)=' _P_<(^ SUBJ)>' } .
PASS (SCHEMATA) = SCHEMATA
(^ PASSIVE) =C +
(^ SUBJ)-->NULL.
```

Following c-structure rules define the constituent structure of a simple sentence with subject. S (sentence) is defined as NP (defined as subject) followed by a verbal complex (VC). Annotations like “(! NUM) = (^ NUM)” will ensure agreement among different constituents of a sentence. VC (verb complex) is defined as a verb (main verb) followed by an optional auxiliary verb. Note that NP is optional here, this optionality will only be used when verb is in passive form and subject becomes NULL as described in PASS template.

```
S --> (NP: (^ SUBJ) =!
( ! GEND) = (^ GEND)
( ! NUM) = (^ NUM) ; )
VC:      ( ! NUM) = (^NUM)
( ! GEND) = (^ GEND)
^=! .
VC --> V:^=! ; (VAUX:^=!) .
```

C-structure and f-structure analysis of simple sentence “CHOkirO dORE” is shown in Figure III-44. It may be noted that NP “CHOkirO” is subcategorized as SUBJ and tense of verb “dORE” is undefined however tense form of verb is aorist which is base form for tense and aspect marking.

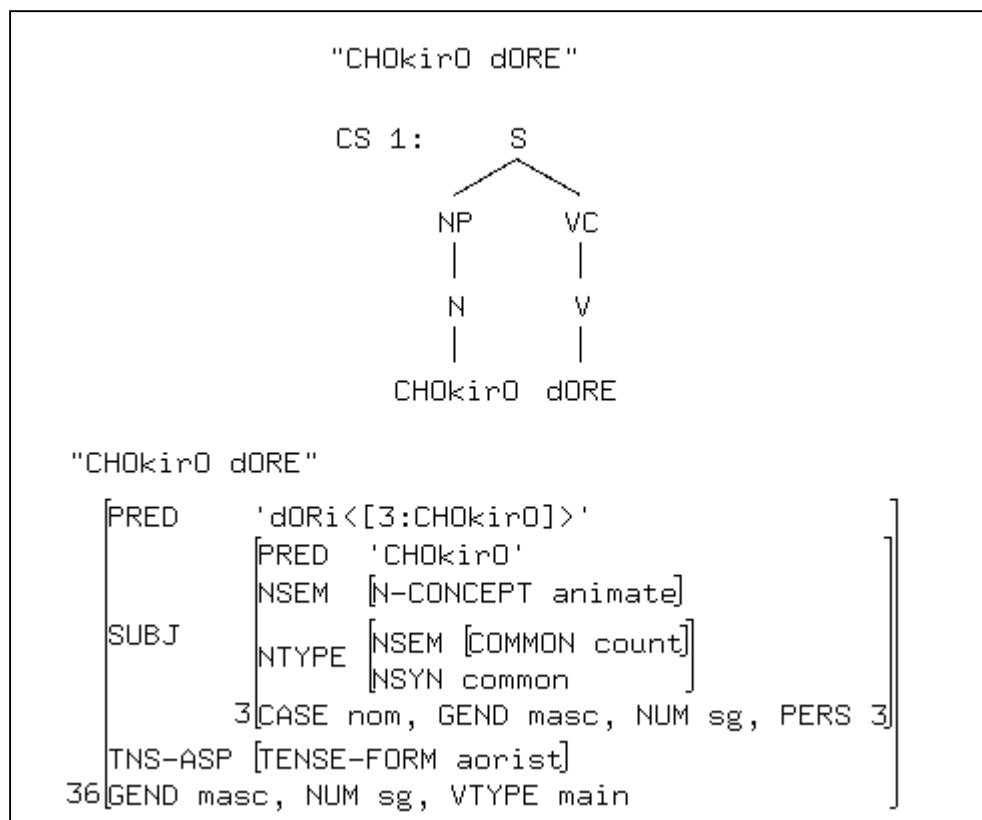


Figure III - 44: LFG Analysis of Sample Sentence “CHOkirO dORE”.

Figure III-45 shows analysis of passive verb form of dORi i.e. dORijE with auxiliary thO.

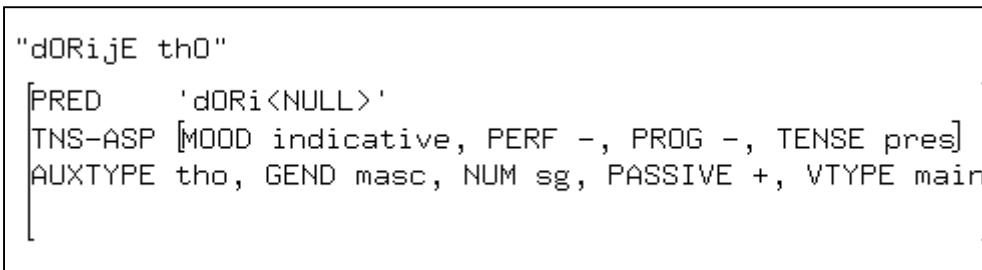


Figure III - 45: F-structure Analysis of Passive Verb “dORijE”.

4.6.2 Object

Usually transitive verbs subcategorize for subject and objects. However, transitive verbs can subcategorize only for subject, or suppressed object (subject) or none of these in case of passive verbs. Following is the lexicon entry of transitive verb “likHu” (write). V-SUBJ-OBJ template defines predicate with SUBJ and OBJ parameters and PASS template defines suppressed object (object is converted to subject in passive formation).

```

likHu  V XLE { @ (V-SUBJ-OBJ %stem) | @ (V-SUBJ %stem) } .

V-SUBJ-OBJ(_P_) ={ @ (PASS (^ PRED)= '_P_<(^ SUBJ) (^
OBJ)>' ) |
                      (^ PRED)= '_P_<(^ SUBJ) (^ OBJ)>' } .

PASS(SCHEMATA) = SCHEMATA
    (^ PASSIVE) =c +
    (^ OBJ) --> (^ SUBJ)
    (^ SUBJ) -->NULL .

```

LFG c-structure rules are also changed and sentence definition will now have an optional KP as an object.

```

S --> (NP: (^ SUBJ)=! (! GEND)=(^ GEND) ; )
        (KP: (^ OBJ)=! { (! CASE)=c acc | (! CASE)=c
nom} )
VC: (! NUM)=(^NUM) (! GEND)=(^ GEND) ^=! .

```

Figure III-46 shows parse tree and f-structure analysis of a sentence with subject and object as transitive verb parameters. As discussed above “likHu” can also have one subject argument only; Figure III-47 shows f-structure analysis of a sentence with subject subcategorization only. It may be noted that in sentence of Figure III-46 “kitAbu” is an object but in Figure III-48 when

passive “likHijE” is used this objects becomes subject and subject becomes NULL. This is due to following rules in PASS template.

(^ OBJ) --> (^ SUBJ)

(^ SUBJ) --> NULL.

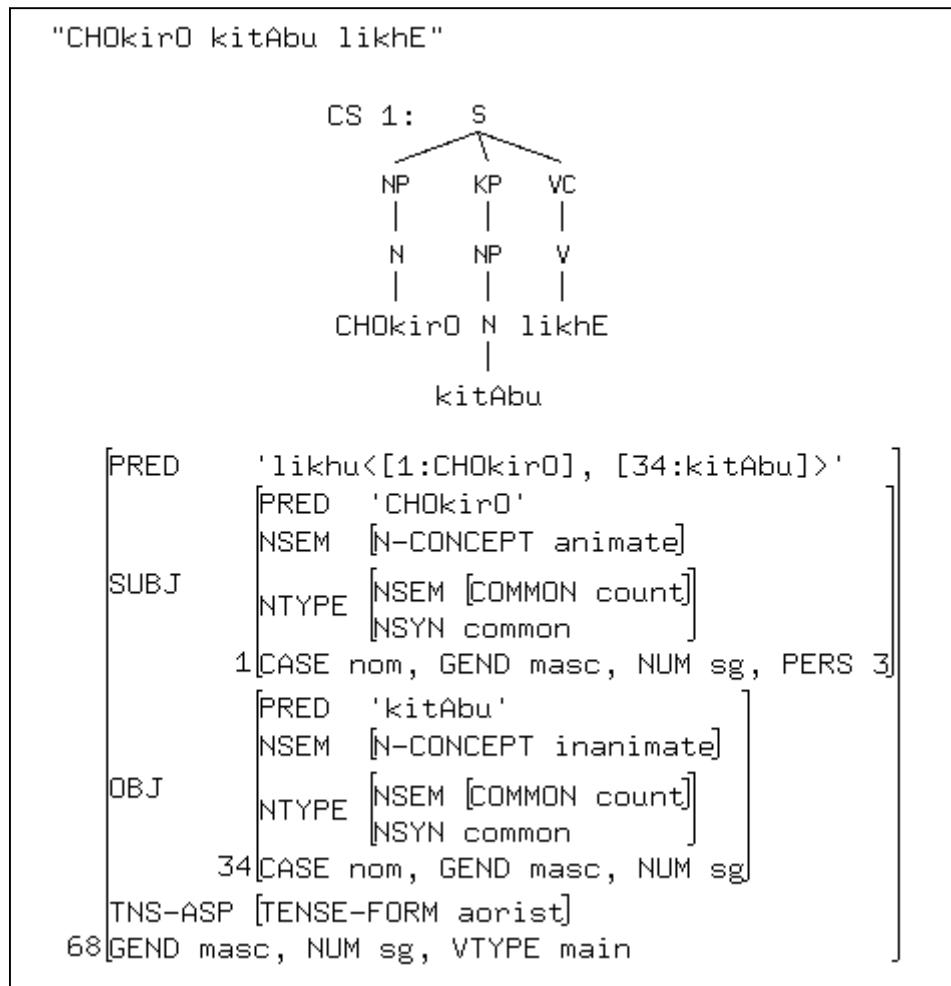


Figure III - 46: Transitive Verb likhu with “SUBJ” and “OBJ” Subcategorization.

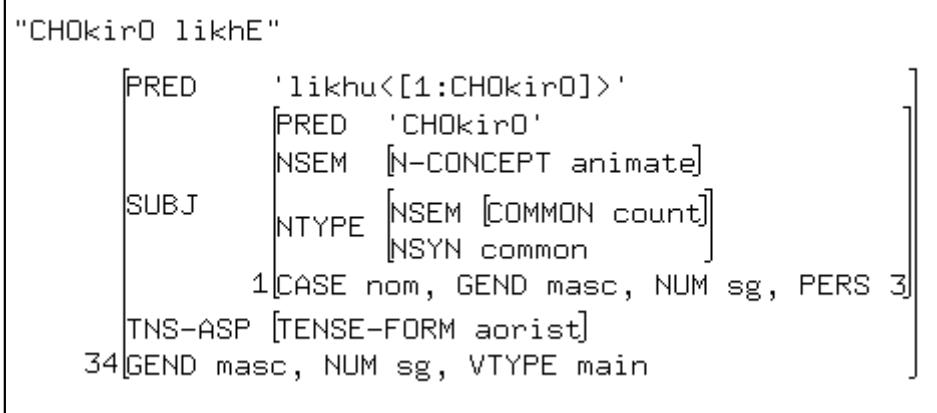


Figure III - 47: Verb "likhu" with SUBJ Subcategorization Only.

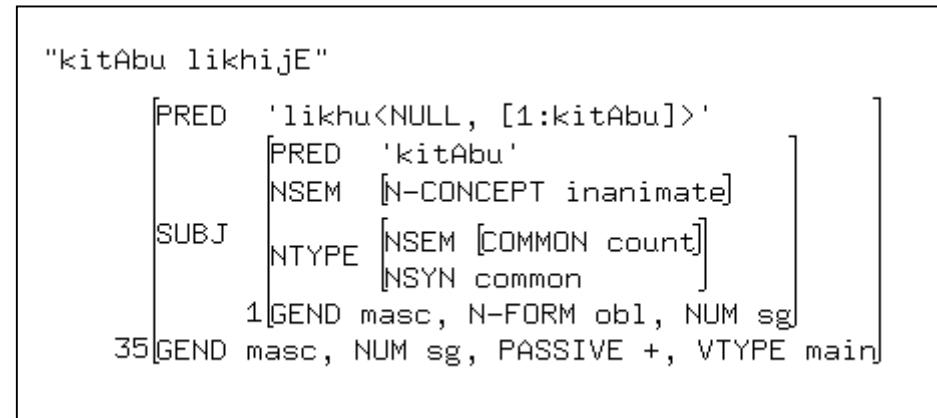


Figure III - 48: Passivization of "likhu" with NULL Subject formation.

4.6.3 Secondary Object (Obj2)

As transitive and causative verbs also subcategorize for indirect objects. Consider following sentence

Ali	CHOkirE=kHE	KHatu	likHE -----	(66)
Ali.Nom.M.Sg	boy.Obl.Sg=Dat	letter.Nom.M.Sg	write.Aorist.Sg.3P	
Ali (may) write a letter to the boy.				

Here "Ali" is a subject in nominative case, "CHOkirE=kHE" is secondary object "CHOkirO" in dative case form, and "KHatu" is direct object; "likHE" therefore

is subcategorized for SUBJ, OBJ2, and OBJ. LFG c-structure rules for such sentences are given below; an extra KP in dative case is introduced to handle secondary objects.

```
S --> (NP: (^ SUBJ)=! (! GEND)=(^ GEND);)
      (KP: (^ OBJ2)=! (! CASE)=c dat)
      (KP: (^ OBJ)=! { (! CASE)=c acc | (! CASE)=c nom} )
      VC: (! NUM)=(^NUM) (! GEND)=(^ GEND) ^=! .
```

V-SUB-OBJ2-OBJ template defines predicates with such subcategorization.

```
V-SUB-OBJ2-OBJ(_P_) =
{@(PASS (^ PRED)=' _P_<(^ SUBJ) (^ OBJ2) (^ OBJ)>' ) |
 (^ PRED)=' _P_<(^ SUBJ) (^ OBJ2) (^ OBJ)>' } .
```

During passive formation of these sentences object becomes subject, secondary object becomes object and subject becomes NULL. PASS template is therefore defined as given below.

```
PASS(SCHEMATA) =SCHEMATA
  (^ PASSIVE) =c +
  (^ OBJ)-->(^ SUBJ)
  (^ OBJ2) --> (^ OBJ)
  (^ SUBJ)-->NULL.
```

Lexicon entry of “likHu” is also updated accordingly.

```
likHu V XLE { @ (V-SUBJ-OBJ %stem) |
 @ (V-SUBJ %stem) |
 @ (V-SUB-OBJ2-OBJ %stem) } .
```

Figure III-49 and Figure III-50 show parse tree and f-structure analysis of a sentence with subject, secondary object and object subcategorization respectively.

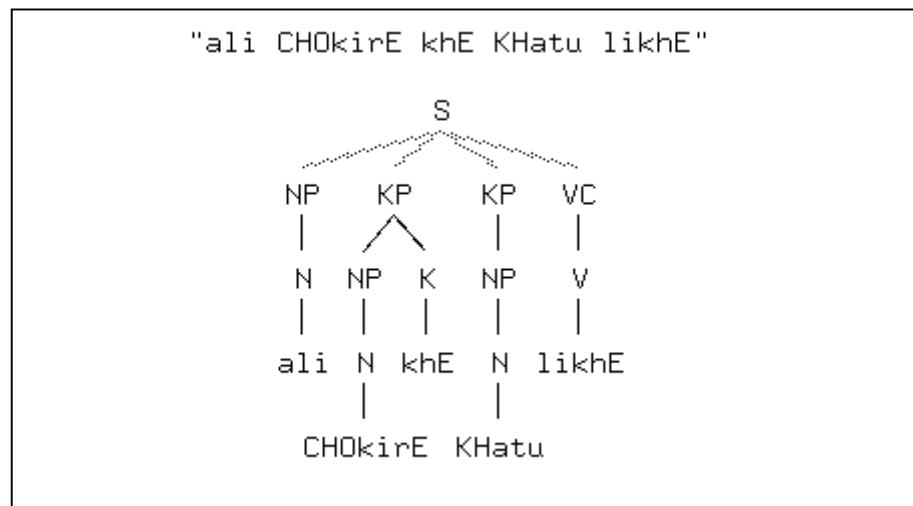


Figure III - 49: Parse Tree for a Sentence with SUBJ, OBJ2, and OBJ.

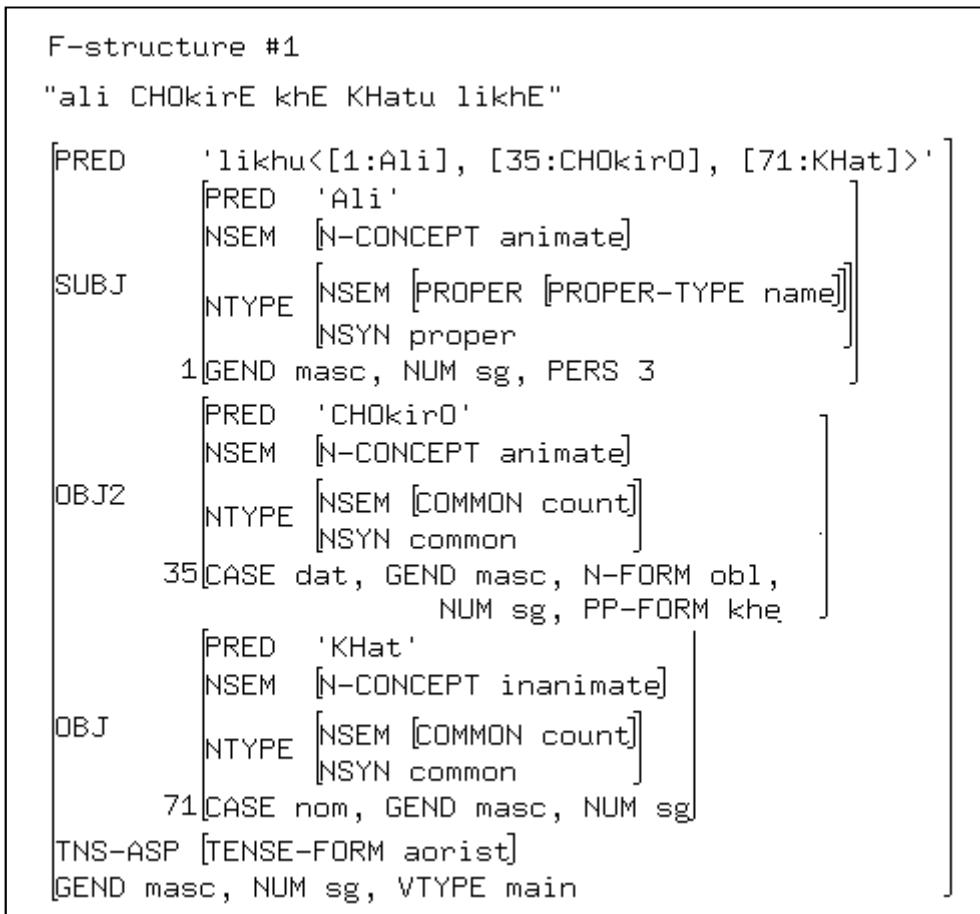


Figure III - 50: F-structure Analysis of a Sentence with SUBJ, OBJ2, and OBJ.

Figure III-51 shows f-structure analysis of passive formation with OBJ2 subcategorization when subject becomes NULL secondary object becomes object and object becomes subject.

F-structure #1					
'KHatu CHOkirE khE likhijE"					
PRED	'likhu<NULL, [35:CHOkir0], [1:KHat]>'				
SUBJ	<table border="0"> <tr> <td>[PRED 'KHAT'</td> <td rowspan="3">NSEM [N-CONCEPT inanimate]</td> </tr> <tr> <td>NSEM [COMMON count]</td> </tr> <tr> <td>NSYN common</td> </tr> </table>	[PRED 'KHAT'	NSEM [N-CONCEPT inanimate]	NSEM [COMMON count]	NSYN common
[PRED 'KHAT'	NSEM [N-CONCEPT inanimate]				
NSEM [COMMON count]					
NSYN common					
OBJ	<table border="0"> <tr> <td>1[CASE nom, GEND masc, NUM sg]</td> <td rowspan="3">NSEM [N-CONCEPT animate]</td> </tr> <tr> <td>[PRED 'CHOkir0'</td> </tr> <tr> <td>NSEM [COMMON count]</td> </tr> </table>	1[CASE nom, GEND masc, NUM sg]	NSEM [N-CONCEPT animate]	[PRED 'CHOkir0'	NSEM [COMMON count]
1[CASE nom, GEND masc, NUM sg]	NSEM [N-CONCEPT animate]				
[PRED 'CHOkir0'					
NSEM [COMMON count]					
	<table border="0"> <tr> <td>35[CASE acc, GEND masc, N-FORM obl,</td> <td rowspan="2">NSYN common</td> </tr> <tr> <td>NUM sg, PP-FORM khe]</td> </tr> </table>	35[CASE acc, GEND masc, N-FORM obl,	NSYN common	NUM sg, PP-FORM khe]	
35[CASE acc, GEND masc, N-FORM obl,	NSYN common				
NUM sg, PP-FORM khe]					
	GEND masc, NUM sg, PASSIVE +, VTTYPE main				

Figure III - 51: Passive Formation with SUBJ, OBJ2, and OBJ.

4.6.4 Oblique

Oblique arguments are marked by agentive or instrumental case in causative verbs. Lexicon entry and V-SUB-OBJ2-OBL-OBJ template definitions are given below. It may be noted that “(^ V-Form2) =c causative” is a necessary condition which requires verb to be in causative form.

```
V-SUB-OBJ2-OBL-OBJ(_P_) =
  (^ PRED)=' _P_<(^ SUBJ) (^ OBJ2) (^ OBL) (^ OBJ)>'.
likhu  V XLE { @ (V-SUBJ-OBJ %stem) | @ (V-SUBJ %stem) |
               @ (V-SUB-OBJ2-OBJ %stem) |
               @ (V-SUB-OBJ2-OBL-OBJ %stem)
               (^ V-Form2)=c causative}.
```

Figure III-52 and Figure III-53 show parse tree and f-structure analysis with agentive oblique argument. V-Form 2 attribute reflects that this is a causative verb.

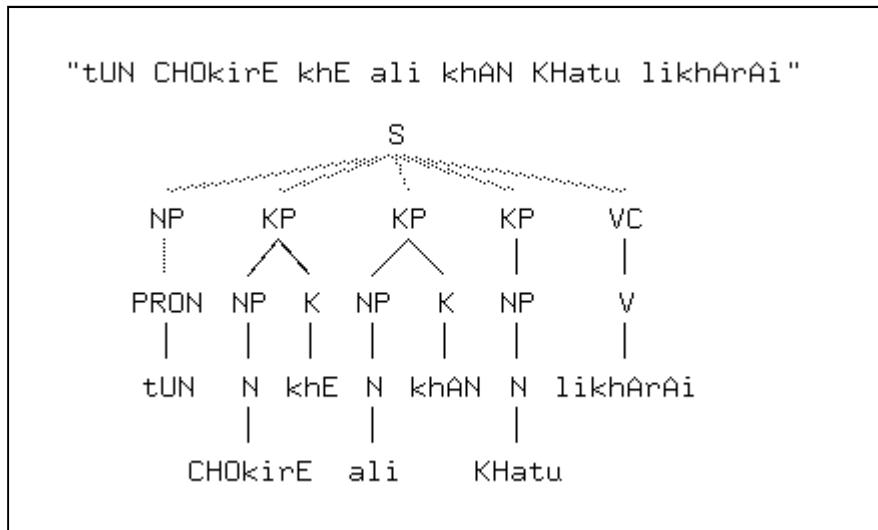


Figure III - 52: Parse Tree of a Sentence with SUBJ, OBJ2, OBL, and OBJ Arguments.

4.6.5 Complement (COMP)

Arguments of verb also include other clauses. Closed complement clause can be embedded within sentences and is subcategorized for verb. At c-structure level this is represented as a complement phrase (CP) and at f-structure level this is known as COMP and is subcategorized for verb. At c-structure level usually a complement phrase is a sentence following the complement marker “ta” (that). Updated lexicon entry and template definition are given below:

```

V-SUBJ-COMP (_P_) = (^ PRED) = '_P_<(^ SUBJ) (^ COMP)>' .
likhu V XLE { @ (V-SUBJ-OBJ %stem) | @ (V-SUBJ %stem) |
@ (V-SUB-OBJ2-OBJ %stem) |
@ (V-SUB-OBJ2-OBI-OBJ %stem)
  
```

```
(^ V-Form2)=c causative|
@ (V-SUBJ-COMP %stem) }.
```

C-structure and f-structure of sample sentence are given in Figure III-54 and Figure III-55 respectively.

```
F-structure #1
'tUN CHOkirE khE ali khAN KHatu likhArAi'

PRED 'likhu<[1:tUN],[35:CHOkir0],[71:Ali],[108:KHAT]>'
  [PRED 'tUN'
    SUBJ NTYPE [NSYN pronoun]
      CASE nom, GEND masc, NUM sg, PERS 2, PRON-TYPE personal ]
  [PRED 'CHOkir0'
    NSEM [N-CONCEPT animate]
    OBJ2 NTYPE [NSEM [COMMON count]
      NSYN common
      CASE dat, GEND masc, N-FORM obl, NUM sg, PP-FORM khe ]
  [PRED 'Ali'
    NSEM [N-CONCEPT animate]
    OBL NTYPE [NSEM [PROPER [PROPER-TYPE name]]
      NSYN proper
      CASE agent, GEND masc, NUM sg, PERS 3, PP-FORM khan ]
  [PRED 'KHAT'
    NSEM [N-CONCEPT inanimate]
    OBJ NTYPE [NSEM [COMMON count]
      NSYN common
      CASE nom, GEND masc, NUM sg
      GEND masc, V-Formimperative, V-Form2 causative, VTYP main ]]
```

Figure III-53: F-structure Analysis of a Sentence with SUBJ, OBJ2, Agentive OBL and OBJ Subcategorization.

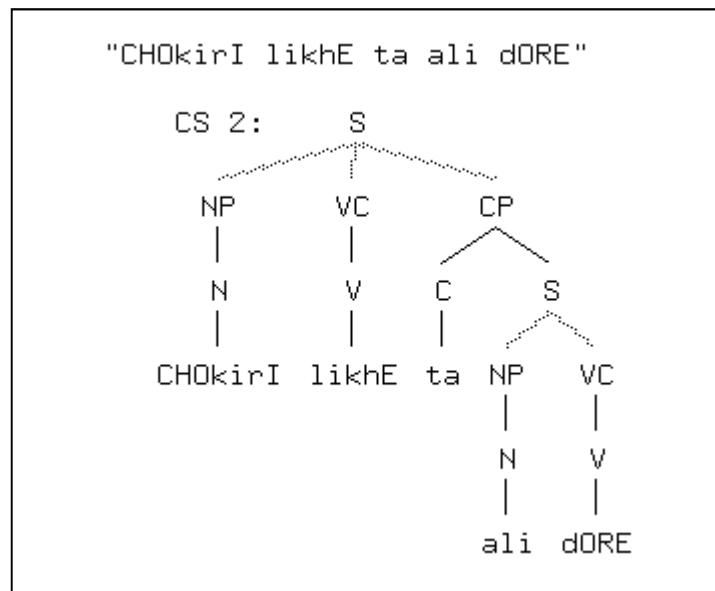


Figure III - 54: Sentence with Embedded Complement Phrase (CP).

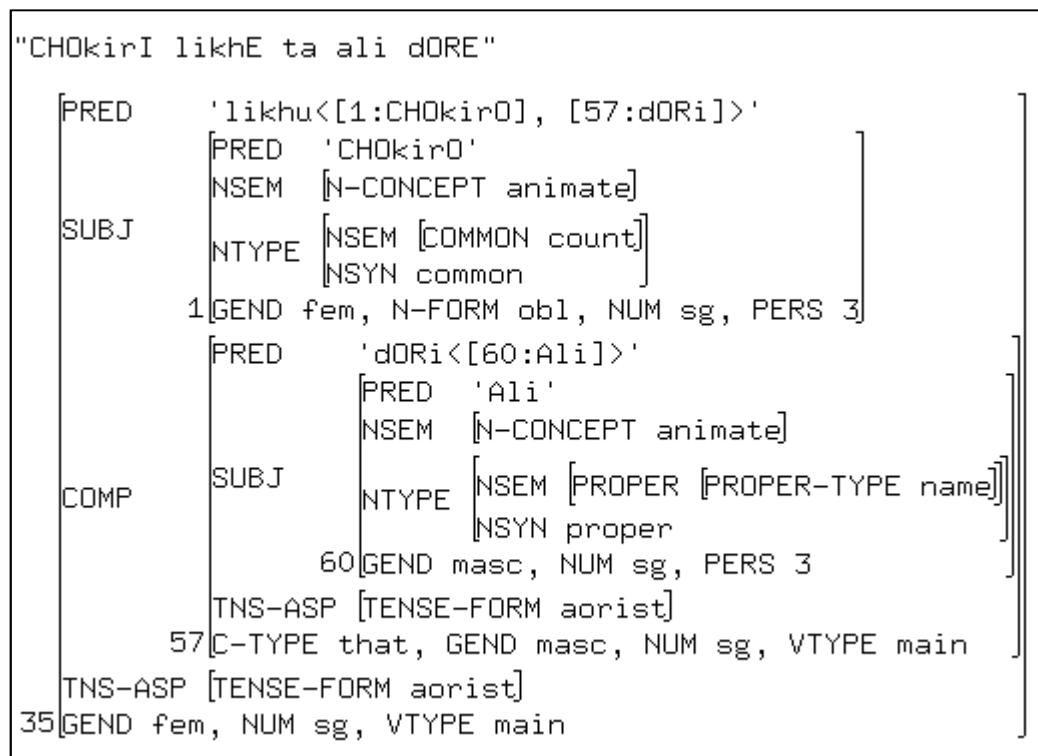


Figure III - 55: F-structure Analysis of a Sentence with COMP Argument.

4.6.6 Open Complement (XCOMP)

Open complement or XCOMP is another embedded clause whose subject is controlled from outside of the clause. XCOMPs can appear with other verbal arguments. In Sindhi XCOMP clauses are represented by NPs or KPs followed by infinitive verb as shown in XCP (XCOMP Phrase) definition below. Following are c-structure rules for a sentence containing SUBJ, OBJ2 and an XCOMP clause.

```

XC --> { KP:(^ SUBJ)=! { (! CASE)=nom | (! N-FORM)=obl } }
          (KP: (^ OBJ2)=! (! CASE)=c dat)
          XCP:(^ XCOMP)=! ;
          VC:^=! .
XCP --> KP: (^ OBJ)=! ; V:^=! (! VFORM)=inf.

```

Following are template definitions and updated lexicon entry.

```

V-SUBJ-XCOMP (_P_) =
    (^ PRED) = '_P_ <(^ SUBJ) (^ XCOMP) > '
    (^ XCOMP SUBJ) = (^ SUBJ) .

V-SUBJ-OBJ2-XCOMP (_P_) =
    (^ PRED) = '_P_ <(^ SUBJ) (^ OBJ2) (^
XCOMP) > '
    (^ XCOMP SUBJ) = (^ SUBJ) .

likhu V XLE { @ (V-SUBJ-OBJ %stem) |
               @ (V-SUBJ %stem) | @ (V-SUB-OBJ2-OBJ
%stem) |
               @ (V-SUB-OBJ2-OBL-OBJ %stem)
               (^ V-Form2)=c causative |
               @ (V-SUBJ-COMP %stem) |
               @ (V-SUBJ-XCOMP %stem) |
               @ (V-SUBJ-OBJ-XCOMP %stem) } .

```

It can be seen in above template definitions that SUBJ of XCOMP is referenced from main sentence i.e. XCOMP SUBJ is actually SUBJ of main sentence. Figure III-56 shows c-structure tree of a sentence with embedded XCP (open complement phrase).

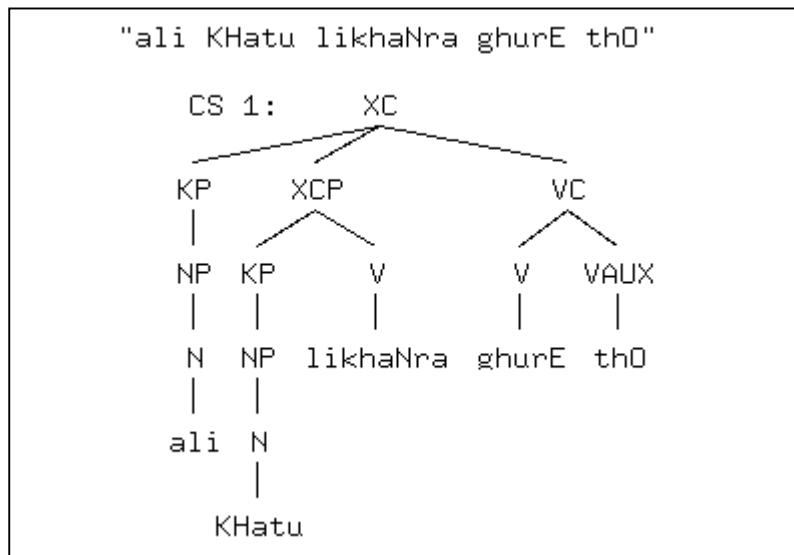


Figure III - 56: Parse Tree of a Sentence with Embedded XCOMP Phrase.

F-structure analysis of same sentence with open complement (XCOMP) subcategorization is shown in Figure III-57. Note that XCOMP has its own OBJ but SUBJ of XCOMP is referenced from PRED of the sentence. Figure III-58 shows another example of XCOMP subcategorization with SUBJ and OBJ2.

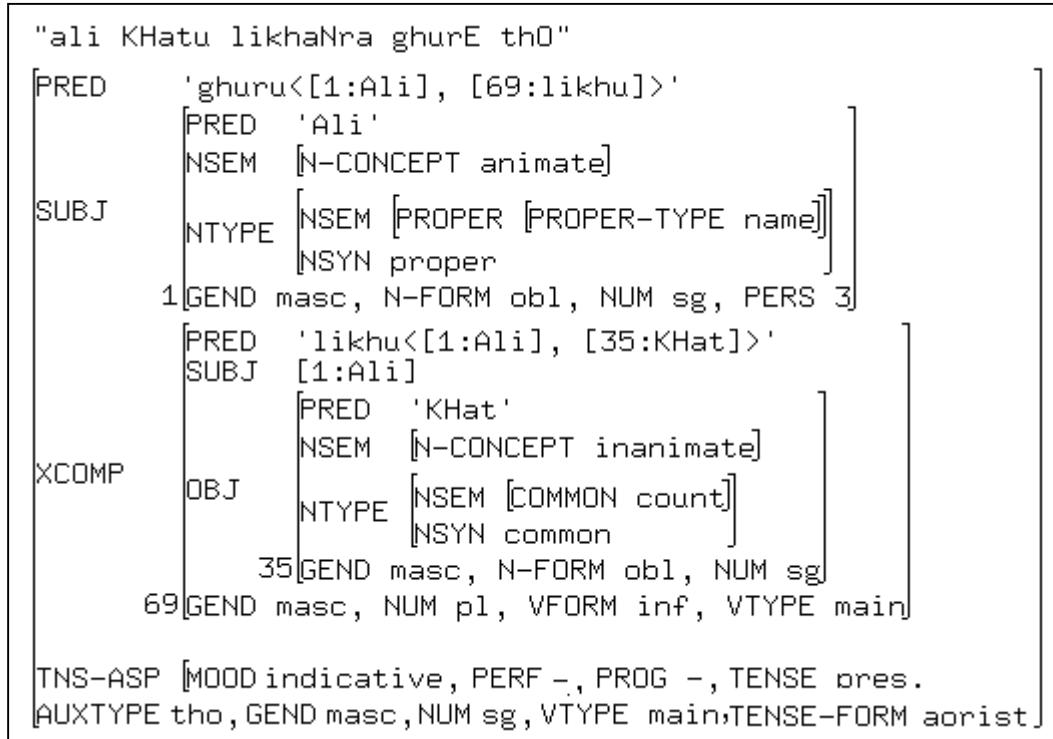


Figure III - 57: F-structure Analysis a Sentence with SUBJ and XCOMP Subcategorization.

"ali CHOkirE khE KHatu likhaNra ghurE tho"

```

PRED      'ghuru<[1:Ali], [35:CHOkir0], [105:likhu]>' ]
          PRED  'Ali'
          NSEM  [N-CONCEPT animate]
SUBJ      NTYPE [NSEM [PROPER [PROPER-TYPE name]]]
          NSYN proper
1[CASE nom, GEND masc, NUM sg, PERS 3 ]
          PRED  'CHOkir0'
          NSEM  [N-CONCEPT animate]
OBJ2      NTYPE [NSEM [COMMON count]]
          NSYN common
35[CASE dat, GEND masc, N-FORM obl, NUM sg, PP-FORM khe
          PRED  'likhu<[1:Ali], [71:KHat]>' ]
          SUBJ [1:Ali]
          PRED  'KHat'
          NSEM  [N-CONCEPT inanimate]
XCOMP     OBJ   NTYPE [NSEM [COMMON count]]
          NSYN common
          71[GEND masc, N-FORM obl, NUM sg]
          105[GEND masc, NUM pl, VFFORM inf, VTYPE main]
TNS-ASP   [MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]
AUXTYPE tho, GEND masc, NUM sg, VTYPE main

```

Figure III - 58: F-structure Analysis a Sentence with SUBJ, XCOMP and OBJ2 Subcategqorization

4.7 Adjunct

Sentence level adjuncts (remember adjuncts are not subcategorized for verbs) include any number of postpositional and adverbial phrases. Following c-structure rules show handling of sentence level adjuncts in Sindhi sentence.

```

S --> (NP: (^ SUBJ)=! (! GEND)=(^ GEND);)
        (KP: (^ OBJ2)=! (! CASE)=c dat)
        (KP: (^ OBL)=! { (! CASE)=c inst | (! CASE)=c
agent})
        (KP: (^ OBJ)=! { (! CASE)=c acc | (! CASE)=c nom})

```

```
( [PPP* : !$ (^ADJUNCT) ; , ADVP*: !$ (^ADJUNCT) ] )
VC: (! NUM)=(^NUM) (! GEND)=(^ GEND) ^!=!;
(CP: (^ COMP)=!) .
```

Adjuncts are placed before verb complex definition; and can either be postpositional or adverbial phrases or their combination with any number of postpositional and adverbial phrases. Shuffle operator “,” ensures shuffling of these phrases i.e. any phrase can occur in any sequence. C-structure tree and f-structure analysis are shown in Figure III-59 and III-60 respectively.

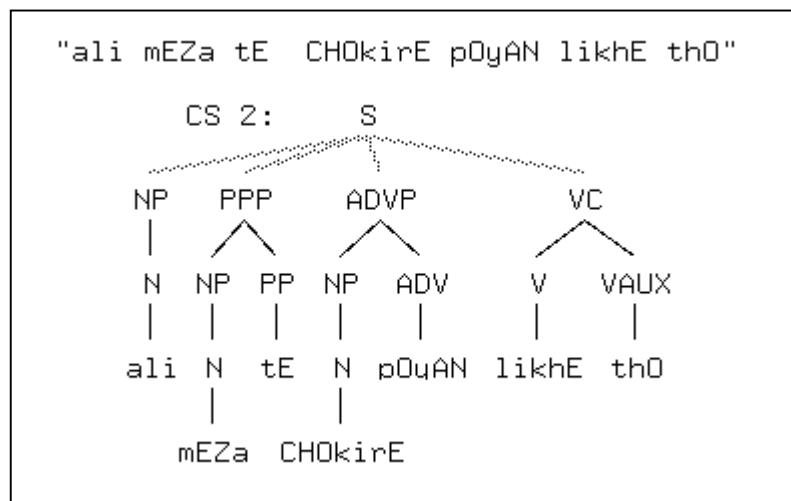


Figure III - 59: Sentence with Adjuncts.

	"ali mEZa tE CHOkirE p0yAN likhE th0"																						
PRED	'likhu<[1:Ali]>'																						
SUBJ	<table border="1"> <tr> <td>PRED</td><td>'Ali'</td></tr> <tr> <td>NSEM</td><td>[N-CONCEPT animate]</td></tr> <tr> <td>NTYPE</td><td> <table border="1"> <tr> <td>NSEM</td><td>[PROPER [PROPER-TYPE name]]</td></tr> <tr> <td>NSYN</td><td>proper</td></tr> </table> </td></tr> <tr> <td>1</td><td>GEND masc, NUM sg, PERS 3</td></tr> </table>	PRED	'Ali'	NSEM	[N-CONCEPT animate]	NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[PROPER [PROPER-TYPE name]]</td></tr> <tr> <td>NSYN</td><td>proper</td></tr> </table>	NSEM	[PROPER [PROPER-TYPE name]]	NSYN	proper	1	GEND masc, NUM sg, PERS 3										
PRED	'Ali'																						
NSEM	[N-CONCEPT animate]																						
NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[PROPER [PROPER-TYPE name]]</td></tr> <tr> <td>NSYN</td><td>proper</td></tr> </table>	NSEM	[PROPER [PROPER-TYPE name]]	NSYN	proper																		
NSEM	[PROPER [PROPER-TYPE name]]																						
NSYN	proper																						
1	GEND masc, NUM sg, PERS 3																						
OBJ	<table border="1"> <tr> <td>PRED</td><td>'tE<[35:mEZa]>'</td></tr> <tr> <td>NSEM</td><td>[N-CONCEPT inanimate]</td></tr> <tr> <td>NTYPE</td><td> <table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table> </td></tr> <tr> <td>35</td><td>GEND fem, N-FORM obl, NUM sg</td></tr> </table>	PRED	'tE<[35:mEZa]>'	NSEM	[N-CONCEPT inanimate]	NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table>	NSEM	[COMMON count]	NSYN	common	35	GEND fem, N-FORM obl, NUM sg										
PRED	'tE<[35:mEZa]>'																						
NSEM	[N-CONCEPT inanimate]																						
NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table>	NSEM	[COMMON count]	NSYN	common																		
NSEM	[COMMON count]																						
NSYN	common																						
35	GEND fem, N-FORM obl, NUM sg																						
ADJUNCT	<table border="1"> <tr> <td>CASE</td><td>loc, PP-FORM on</td></tr> <tr> <td>PRED</td><td>'p0yAN<[72:CHOkir0]>'</td></tr> <tr> <td>NSEM</td><td>[N-CONCEPT animate]</td></tr> <tr> <td>OBJ</td><td> <table border="1"> <tr> <td>NTYPE</td><td> <table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table> </td></tr> <tr> <td>72</td><td>GEND masc, N-FORM obl, NUM sg</td></tr> <tr> <td>105</td><td>ATYPE space</td></tr> </table> </td></tr> <tr> <td>TNS-ASP</td><td>[MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]</td></tr> <tr> <td>112</td><td>AUXTYPE tho, GEND masc, NUM sg, VTYPE main</td></tr> </table>	CASE	loc, PP-FORM on	PRED	'p0yAN<[72:CHOkir0]>'	NSEM	[N-CONCEPT animate]	OBJ	<table border="1"> <tr> <td>NTYPE</td><td> <table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table> </td></tr> <tr> <td>72</td><td>GEND masc, N-FORM obl, NUM sg</td></tr> <tr> <td>105</td><td>ATYPE space</td></tr> </table>	NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table>	NSEM	[COMMON count]	NSYN	common	72	GEND masc, N-FORM obl, NUM sg	105	ATYPE space	TNS-ASP	[MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]	112	AUXTYPE tho, GEND masc, NUM sg, VTYPE main
CASE	loc, PP-FORM on																						
PRED	'p0yAN<[72:CHOkir0]>'																						
NSEM	[N-CONCEPT animate]																						
OBJ	<table border="1"> <tr> <td>NTYPE</td><td> <table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table> </td></tr> <tr> <td>72</td><td>GEND masc, N-FORM obl, NUM sg</td></tr> <tr> <td>105</td><td>ATYPE space</td></tr> </table>	NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table>	NSEM	[COMMON count]	NSYN	common	72	GEND masc, N-FORM obl, NUM sg	105	ATYPE space												
NTYPE	<table border="1"> <tr> <td>NSEM</td><td>[COMMON count]</td></tr> <tr> <td>NSYN</td><td>common</td></tr> </table>	NSEM	[COMMON count]	NSYN	common																		
NSEM	[COMMON count]																						
NSYN	common																						
72	GEND masc, N-FORM obl, NUM sg																						
105	ATYPE space																						
TNS-ASP	[MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]																						
112	AUXTYPE tho, GEND masc, NUM sg, VTYPE main																						

Figure III - 60: F-structure Analysis a Sentence with XCOMP Subcategorization.

4.8 Open Adjunct (XAdjunct)

Open adjuncts in a sentence are clauses whose subject is controlled from outside of the clause (like XCOMP) and are not subcategorized for verbs (unlike XCOMP). In Sindhi XADJUNCT follows either conjunctive participle or oblique habitual form of verb. Figures III-61 and III-62 show parse tree and f-structure analysis of a sentence with XADJUNCT following conjunctive participle. Following are c-structure rules for sentences with XADJUNCT, only the verb and XADJUNCT definitions are shown here rest of the sentence rules will remain as they are discussed previously.

```

S --> ...
V: (! NUM)=(^NUM)
    (! GEND)=(^ GEND)
    (^ PTCPL-TYPE)=conjunctive
    (^XADJUNCT SUBJ)=(^ SUBJ)
    ^=! ;
S: (^ XADJUNCT)=! .

```

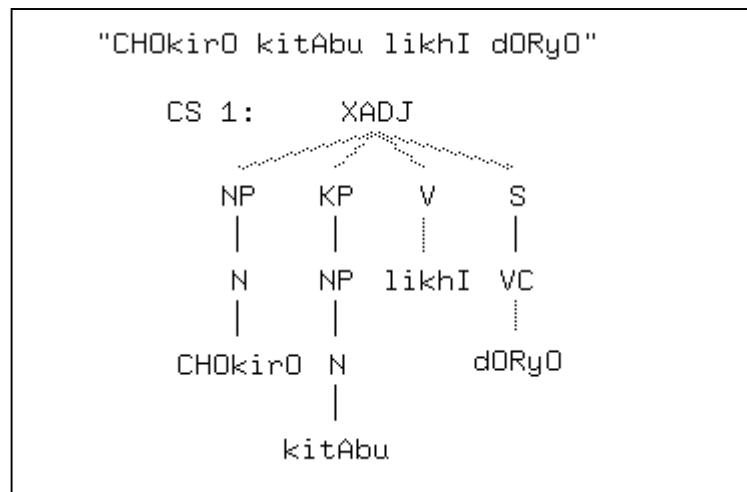


Figure III - 61: Sentence with XADJUNCT.

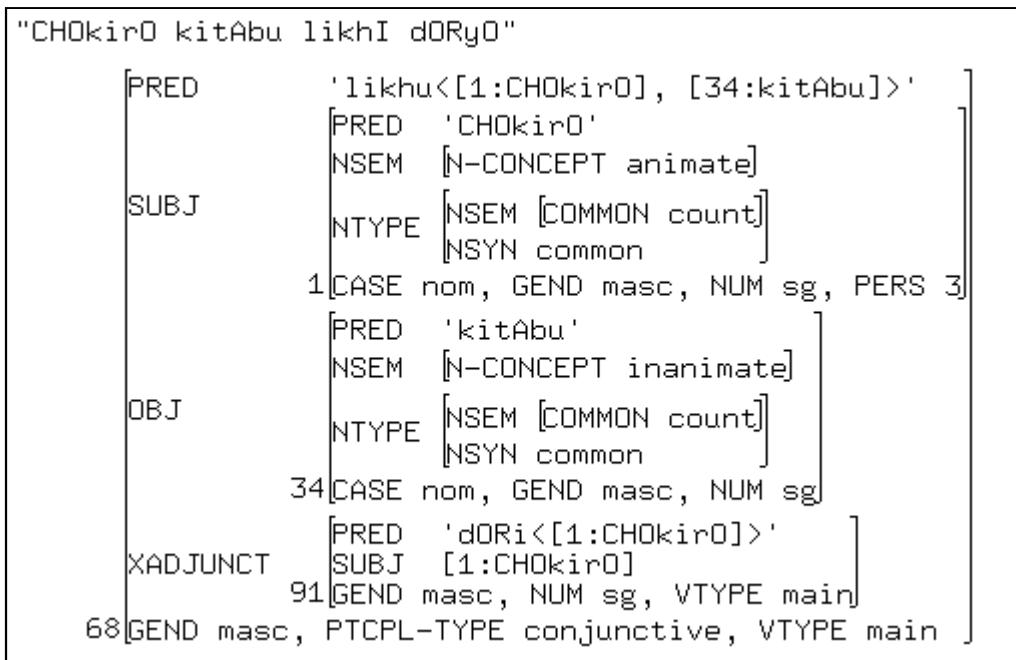


Figure III - 62: F-structure Analysis of a Sentence with XADJUNCT.

4.9 Relative Clause and NP with Relative Clause

Relative clause starts with relative pronoun (subcategorized as subject) and has a verb possibly with other arguments along with SUBJ. However relative clause cannot stand alone as a sentence and is a dependent clause. Following c-structure rules show definition of a relative clause. Notice that after relative pronoun the sentence structure is same like an ordinary sentence.

```
CPRel --> PRON: (! PRON-TYPE)=relative
                  (^ TOPIC-REL)=!
                  (^ TOPIC-REL TOPIC-TYPE)=relative-clause
                  (^ TOPIC-REL)=(^ SUBJ);
(KP: (^ OBJ2)=! (! CASE)=c dat)
(KP: (^ OBL)=! { (! CASE)=c inst | (! CASE)=c agent})
(KP: (^ OBJ)=! { (! CASE)=c acc | (! CASE)=c nom})
([PPP*: !$(^ADJUNCT); , ADVP*: !$(^ADJUNCT)]) )
VC: (! NUM)=(^NUM) (! GEND)=(^ GEND) ^=! .
```

A relative clause parse tree and f-structure analysis is shown in figures III-63 and III-64 respectively.

Noun phrase can also have relative clause as its adjunct. A noun phrase definition with relative clause is given below:

```
NPRel --> NP:!^; CPRel+: !$(^ ADJUNCT)
                  (! ADJUNCT-TYPE)=relative.
```

NP followed by one or more number of relative clause adjuncts defines the NPRel (NP with relative clause). Figure III-65 shows f-structure analysis of such noun phrase.

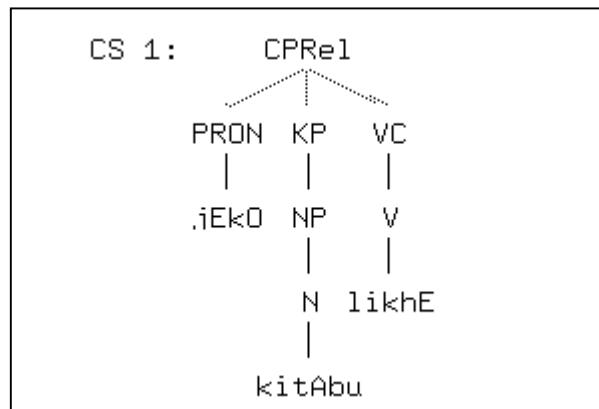


Figure III - 63: Parse Tree of a Relative Clause

```

"jEk0 kitAbu likhE"
[PRED      'likhu<[1:jEk0], [28:kitAbu]>' ]
|          [PRED      'jEk0' ]
|          |          [NTYPE [NSYN pronoun] ]
|          |          [1CASE nom, GEND masc, NUM sg,
|          |          |          [PERS 3, PRON-TYPE relative,
|          |          |          |          .TOPIC-TYPE relative-clause]
|          |          |          [PRED      'kitAbu' ]
|          |          |          |          [NSEM [N-CONCEPT inanimate] ]
|          |          |          |          [NTYPE [NSEM [COMMON count] ]
|          |          |          |          |          [NSYN common]
|          |          |          |          [28CASE nom, GEND masc, NUM sg]
|          |          |          [TNS-ASP [TENSE-FORM aorist]
|          |          |          [TOPIC-REL [1:jEk0]
|          |          |          [NUM sg, VTYPE main
|          ]
  
```

Figure III - 64: F-structure Analysis of a Relative Clause.

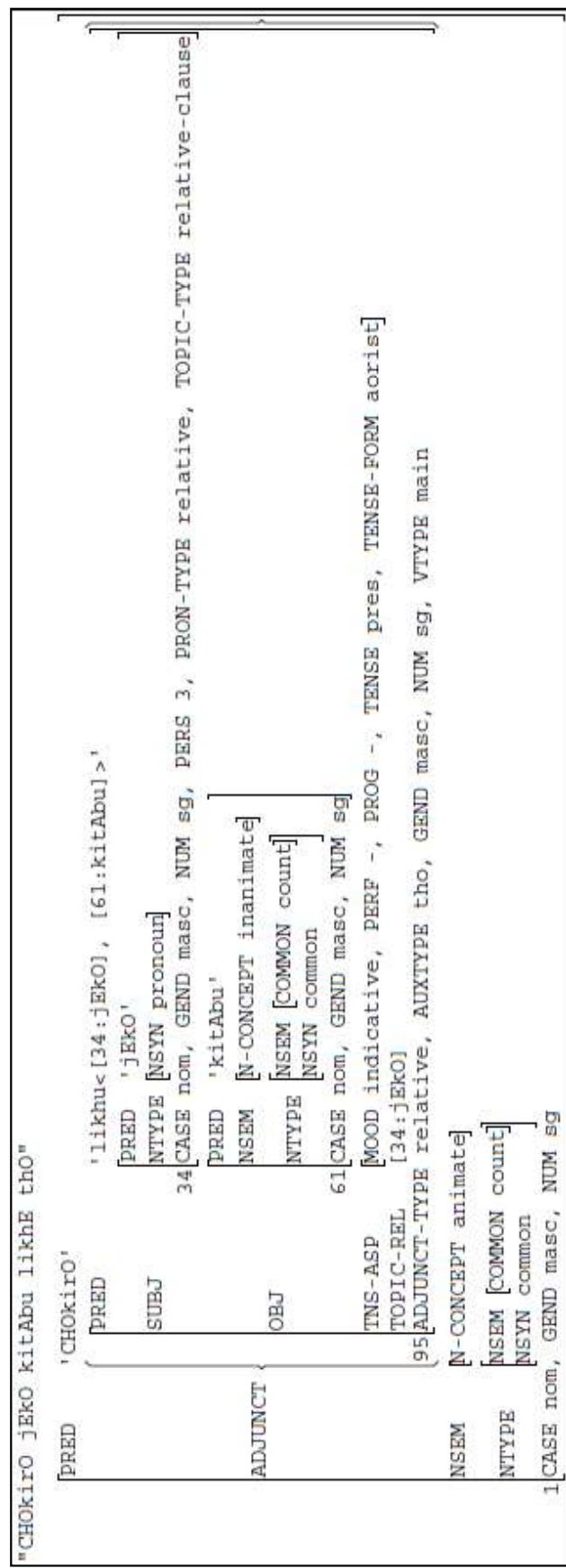


Figure III - 65: F-structure Analysis of an NP with Relative Clause.

4.10 Pronominal Suffixes

As discussed in Section 1, pronominal suffixes are treated as special lexical entries in lexicon. For example, consider transitive verb “likhu” (write); when appears with 1st person pronominal suffix “-iyami” becomes “likh-iyami” (I wrote). Morphological analysis of “likhiyami” is given below:

```
{ likhiyami "+Token" | likhu "+Verb" "+Psx" "+SSg" "+S1P"
  "+SMF" "+SObl" "+Sg" "+PastPart" }
```

Above morphological analysis says that “likhiyami” is a morphological form of root “likhu”. +Psx attribute says that this is a pronominal suffixed form. The tag pattern “+Sxxx” represent different attributes of subject reflected by pronominal suffix. +PastPart tag says that verb form is a past participle. Syntax analysis of “likhiyami” is shown in Figure III-66. Different attributes of verb “likhu” in f-structure are extracted from morphological tags given in above morphological analysis. Pronominal suffixation may cause a complete sentence replaced with single word form with all its verbal and nominal elements. Syntax analysis therefore need to extract / reconstruct this information from morphology. In this case the sentence “mUN likhiyO” (I wrote) is replaced by “likhiyami”. This reconstruction can be seen in verbal subcategorization of “likhu” where the SUBJ argument contains the value ‘pro’ which represents a pronoun with gender, noun form, number and person attributes (feminine, oblique, singular, 1st person). Oblique singular 1st person pronoun in Sindhi is “mUN” which can either be feminine or masculine. As the verb is a past participle therefore its aspect is perfective.

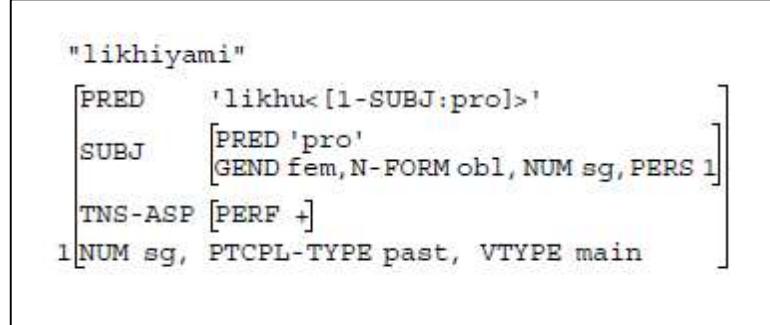


Figure III - 66: F-structure Analysis of verb "likhu" with Pronominal Suffix.

CHAPTER IV

RESULTS

Morphology and syntax implementation coverage discussed in previous chapter includes inflectional morphology of nouns, pronouns, and adjectives with number, gender and case inflections, verbs with number, gender, tense, aspect, mood and causative inflections, adverbs, auxiliaries, modals, and copula verbs. Conjunction, interjection and postposition morphology is implemented as a full form lexicon as a part of LFG grammar development. Different elements of nominal and verbal syntax are implemented in LFG as discussed in Chapter III.

Evaluation of developed morphology is done in terms of precision, recall, coverage and average ambiguity. To evaluate the developed LFG two different test suites are created. Testsuite-1 is a collection of 10 different files of handcrafted sentences and Testsuite-2 is a collection of two different files containing sentences from two textbooks of class 1. Following sections discuss the finite state morphology and lexical function grammar evaluation details along-with test suites, grammatical features covered and tested against these test suites, results, and morphology and syntax coverage.

1. EVALUATION OF FINITE STATE MORPHOLOGY

Coverage of morphology of different POS classes implemented in Xerox Finite State Transducers is shown in Table IV-1 which includes number of stem/root forms and number of morphological forms / surface forms. Total

31 different inflectional categories of nouns are identified and implemented with number, gender, and case paradigms. In the same way 30 different categories of verbs are identified including 21 main verb categories and 9 auxiliary/copula verb categories.

Table IV-1: Stems and Surface Forms Coverage.

Word Class	Stems	Surface Forms
Nouns	418	2086
Verbs	136	7339
Pronouns	129	283
Adjectives	90	508
Adverbs	44	45
Conjunctions	18	18
Postpositions	17	38
Interjections	10	10
Total	862	10327

Different word classes with their implemented number of inflectional categories and average inflections per stem are shown in Table IV-2. Verbs in Sindhi have most average inflections per stem. The next in this list is adjectives entry which is far behind in number (5.64 vs 53.96).

Table IV-2: Inflectional Categories and Average Inflections

Word Class	Number of Inf. Categories	Average Inflections /Stem	Overall Average Inflections
Nouns	31	4.99	
Verbs	30	53.96	
Pronouns	7	2.19	
Adjectives	12	5.64	
Adverbs	2	1.02	
Conjunctions	1	1.00	
Postpositions	2	2.24	
Interjections	1	1.00	
Total	86	-	11.80

A total of 86 inflectional categories of different word classes are

identified and implemented. The overall average inflections per stem observed are 11.8. The developed morphological lexicon is tested against 9050 words corpus. Two different sets of sentences were considered. First set contains 1390 sentences with 7809 words covering different morphological forms of various word classes. Second set contains 258 sentences with 1241 words from two text books of Sindhi class one. Morphological analysis is done by integrating the developed finite state transducers in XLE. Sample morphological analysis of a sentence is given below:

- (1) kAlh asIN mEIO gHumaNra vayAsIN
 yesterday we fair visit.Inf went.1P.Sg
 'We went to see/visit fair yesterday'

```
kAlh {"+Adverb" "+Temp"}  

asIN {"+Pron" "+1P" "+Pl" {"+Masc" | "+Fem"} "+Nominative"}  

mEIO {"+Noun" "+Common" "+Count" "+Inanimate" "+Masc" "+Sg"  

"+Nominative"}  

{gHumu "+Verb" "+Inf"}  

{vaNGa "+Verb" "+SPl" "+S1P" "+SMF" "+SObl" "+Pl" "+PastPart"}
```

In morphological analysis of sentence 1 shown above it can be seen that "kAlh" is identified as a temporal adverb by "+Adverb" and "+Temp" tags, "asIN" is a first person (+1P) plural (+Pl) pronoun (+Pron) with masculine or feminine (+Masc | +Fem) gender in nominative case (+Nominative). In the same way "mEIO" is a common (+Common), inanimate (+Inanimate), count (+Count) noun with masculine gender, singular number and nominative case. "gHumaNra" is an infinitive form (+Inf) of verb root "gHumu" and "vayAsIN" is inflection of verb root "vaNGa" with plural past participle form along-with pronominal suffix attributes starting with "+S" which represents plural, 1st person, masculine or feminine subject in oblique form. First three words of the above sentence are root/stem words and morphological analysis only attaches

the lexical attributes, however, last two words are inflected and their attributes are given accordingly. Morphological analyses with more inflected words are given below in sentences 2 and 3.

- (2) suhiNrA CHOkirA aJi asAN DANhN kitAba paRhaNra- AyA
beautiful.PI.M boys today we to books read.Inf- came.PI.M
"Handsome boys came to us today to read books"

```
{suhiNrO "+Adjective" "+Nominative" "+Masc" "+Pl"}  

{ CHOkirO "+Noun" "+Common" "+Count" "+Animate" {"+Sg" "+Masc"  

"+Vocative" "+2P" textpipe "+Pl" "+Masc" "+Nominative"}}  

aJi {"+Adverb" "+Temp"}  

{asIN "+Pron" "+1P" "+Pl" {"+Masc" | "+Fem"} "+Oblique"}  

DANhN {"+Adverb" "+Space"}  

{kitAba kitAbu "+Noun" "+Common" "+Count" "+Inanimate" "+Masc" "+Pl"  

"+Nominative"}  

{paRhu "+Verb" "+Inf"}  

{acha "+Verb" "+PastPart" "+Masc" "+Pl"}
```

- (3) sabHAnI CHOkirayani kHE SuTHA kapaRA pAtala huA
All girls.Obl Dative Descent Cloths Wear.PastPart Were
All girls were wearing descent clothes.

```
sabHAnI {"+Adjective" "+Oblique" "+Pl"}  

{CHOkir "+Noun" "+Common" "+Count" "+Animate" "+Pl" "+Fem"  

"+Oblique"}  

{kHE "+PostPos" "khe"}  

{suTHO "+Adjective" "+Nominative" "+Masc" "+Pl"}  

{kapaRO "+Noun" "+Common" "+Count" "+Inanimate" "+Masc" "+Pl"  

"+Nominative"}  

{pAi "+Verb" "+PastPart"}  

{AhE {"+Aux" "+AhE" | "+Cop"} "+Past" "+Masc" "+Pl" {"+2P" | "+3P"}}
```

Sentences 2 and 3 cover noun, pronoun, adjective, postposition, verb, auxiliary verb and adverb classes. Apart from adverb and postposition classes all others are inflected forms of their respective root/stem words.

Precision, recall, f-measure, coverage and ambiguity results are shown in Table IV-3. A comparison with Apertium (discussed in Chapter II) is given in Table IV-4.

Table IV-3: Results of Morphological Analysis

Word Class	Precision %	Recall %	F1 Measure %	Coverage %	Ambiguity
Nouns	97.85	96.07	96.95	96.37	1.88
Verbs	98.03	96.89	97.46	98.83	1.57
Adjectives	96.83	93.85	95.31	88.73	1.14
Adverbs	100.00	93.33	96.55	90.00	1.07
Pronouns	95.00	95.00	95.00	88.89	2.12
Postpositions	100.00	100.00	100.00	87.50	1.25
Interjections	100.00	100.00	100.00	80.00	1.00
Conjunctions	100.00	100.00	100.00	98.50	1.00
(Overall)	97.80	96.08	96.93	91.1025	1.65

Table IV-4: Comparison with Apertium Implementation

	Precision %	Recall %	F1-measure %	Avg. Inf./Stem	Avg. Ambiguity
Xerox	97.8	96.08	96.93	11.98	1.65
Apertium	97.68	97.52	97.5	NA	3.3

2. TEST SUITES AND LFG EVALUATION

As discussed above, two different test suites are created. A handcrafted test suite with 10 files and a test suite of class 1 text books. Following sections discuss each test suite in detail.

2.1 Description of Test Suite-1

Test suite 1 is a collection 10 different files each of which contains sentences with grammatical features to be tested against developed grammar. Details of every file are discussed below along-with an example sentence c-structure and f-structure analysis results.

2.1.1 Test File 1

Test file 1 contains total of 106 sentences which cover different grammatical features with PREDLINK and copula verb analysis. Different paradigms of number and gender agreements and pronominal subjects are covered. C-structure and f-structure analysis of a sentence from Test file 1 is shown in Figure IV-1 and Figure IV-2 respectively.

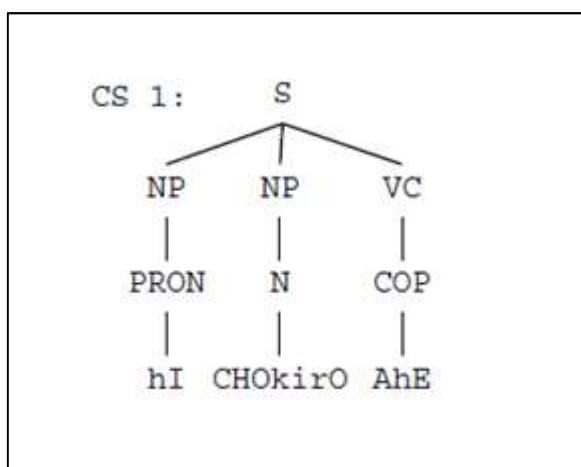


Figure IV - 1: C-structure of a sentence with PREDLINK and Copula Verb.

"hI CHOkirO AhE"	
PRED	'be<[1:hI], [37:CHOkirO]>'
SUBJ	[PRED 'hI' 108 [NTYPE [NSYN pronoun] 110 CASE nom, GEND masc, NUM sg, PERS 3, PRON-TYPE personal]
PREDLINK	[PRED 'CHOkirO' NSEM [N-Concept animate] 37 [NTYPE [NSEM [COMMON count] 126 [NSYN common 128 CASE nom, GEND masc, NUM sg]
69	
135	
137	TNS-ASP [TENSE aorist]
138	VTYPE copular

Figure IV - 2: F-structure Analysis of a sentence with PREDLINK and Copula Verb.

2.1.2 Test File 2

Grammatical features covered in Test file 2 include adjective PREDLINK, adjective coordination, and locative postpositions, along-with all the features of Test file 1. Figure IV-3 shows c-structure tree of a sentence from Test file 2 with coordinated adjectives used as PREDLINK. F-structure analysis of same sentence is shown in Figure IV-4. File consists of 73 sentences.

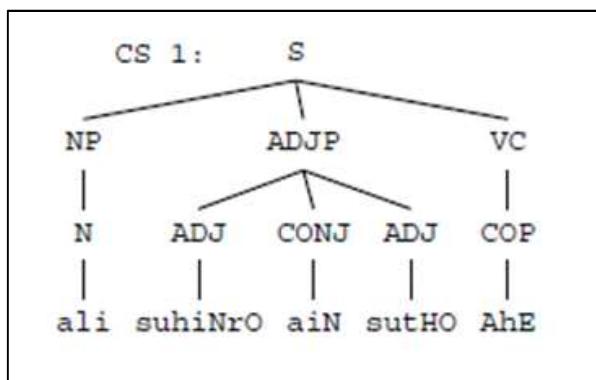


Figure IV - 3: C-structure of a sentence with Coordinated Adjectives PREDLINK.

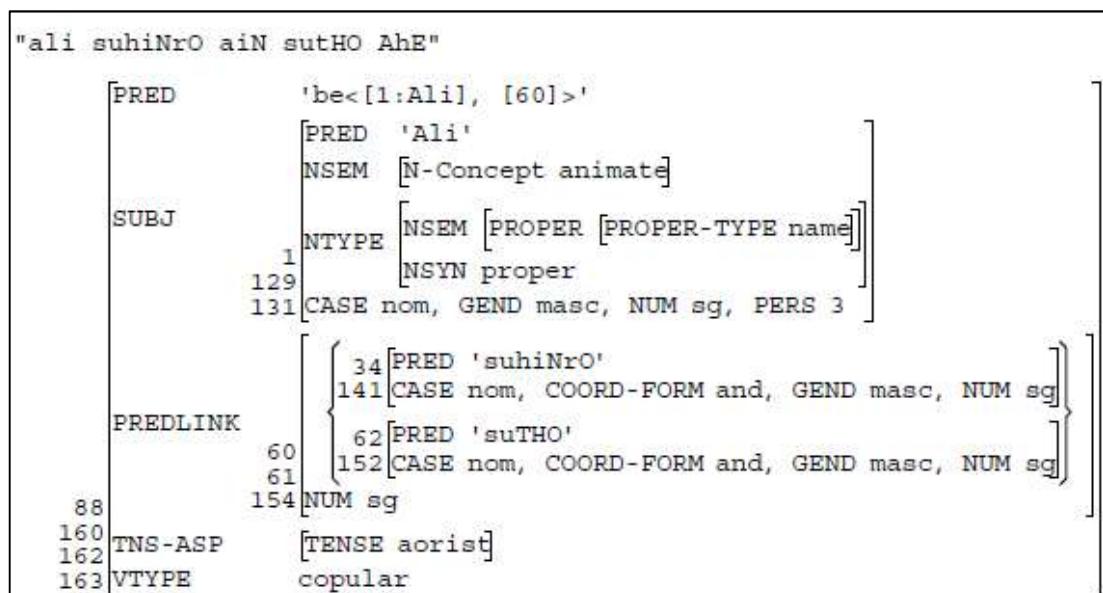


Figure IV - 4: F-structure of a sentence with Coordinated Adjectives PREDLINK.

2.1.3 Test File 3

A total of 81 sentences in Test file 3 cover various grammatical features including number, gender, case and tense agreement, 1st, 2nd and 3rd person pronouns, oblique forms and compound noun phrases with pronoun, noun and adjective coordination. An example sentence analysis from Test file 3 with compound NP construction, present tense, and imperfective continuous aspect is shown in Figure IV-5 and Figure IV-6 respectively.

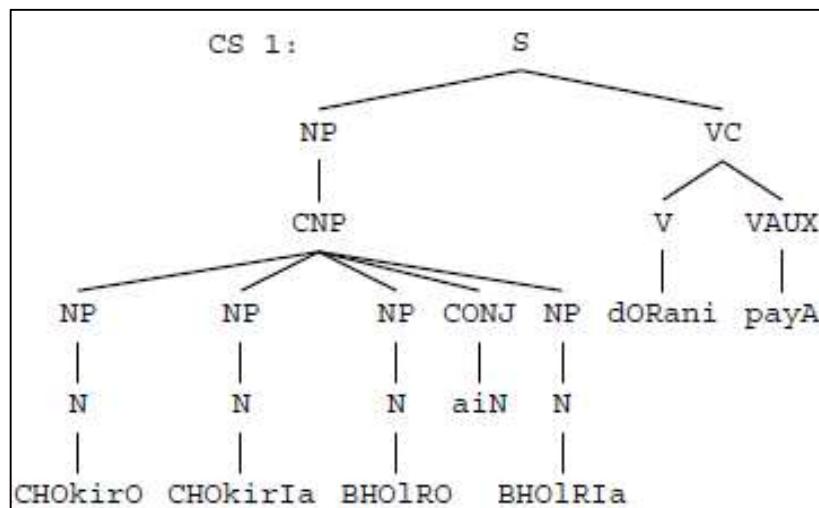


Figure IV - 5: C-structure of a sentence with Compound NP Construction.

'CHOkirO CHOkirIa BHOLRO aiN BHOLRIa dORani payA"	
PRED	'dORi<[97]>' {
	PRED 'CHOkirO' NSEM [N-Concept animate] 1 NTTYPE [NSEM [COMMON count]] 194 NSYN common 196 CASE nom, COORD-FORM and, GEND masc, NUM sg { PRED 'CHOkirO' NSEM [N-Concept animate] 33 NTTYPE [NSEM [COMMON count]] 213 NSYN common 215 CASE nom, COORD-FORM and, GEND fem, NUM sg { PRED 'BHOLRO' NSEM [N-Concept animate] 65 NTTYPE [NSEM [COMMON count]] 230 NSYN common 232 CASE nom, COORD-FORM and, GEND masc, NUM sg { PRED 'BHOLRO' NSEM [N-Concept animate] 97 NTTYPE [NSEM [COMMON count]] 98 NSYN common 252 CASE nom, COORD-FORM and, GEND fem, NUM sg { 153 256 NUM pl, PERS 3 276 257 TNS-ASP [MOOD indicative, PROG +, TENSE pres, TENSE-FORM aorist] 131 266 AUXTYPE payo, GEND masc, NUM pl, VTYPE main
SUBJ	}

Figure IV - 6: F-structure Analysis of a Sentence with Compound NP Construction.

2.1.4 Test File 4

Along-with tense, aspect, coordination Test file 4 covers passive forms of verbs with total number of 32 sentences. Figure IV-7 and IV-8 show c-structure and f-structure of a sentence with passive verb from Test file 4.

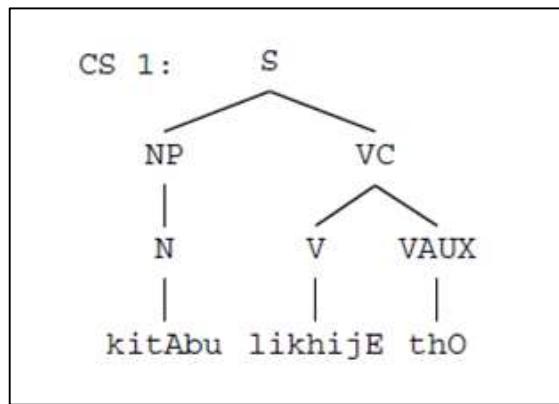


Figure IV - 7: C-structure of a sentence with Passive Verb.

"kitAbu likhijE tho"													
PRED	'likhu<[1:kitAbu]>'												
SUBJ	<table border="0"> <tr> <td>[PRED 'kitAbu'</td> <td></td> </tr> <tr> <td>NSEM [N-Concept inanimate]</td> <td></td> </tr> <tr> <td>NTYPE</td> <td> <table border="0"> <tr> <td>[NSEM [COMMON count]]</td> <td></td> </tr> <tr> <td>NSYN common</td> <td></td> </tr> </table> </td> </tr> <tr> <td>1</td> <td>CASE nom, GEND masc, N-FORM obl, NUM sg</td> </tr> </table>	[PRED 'kitAbu'		NSEM [N-Concept inanimate]		NTYPE	<table border="0"> <tr> <td>[NSEM [COMMON count]]</td> <td></td> </tr> <tr> <td>NSYN common</td> <td></td> </tr> </table>	[NSEM [COMMON count]]		NSYN common		1	CASE nom, GEND masc, N-FORM obl, NUM sg
[PRED 'kitAbu'													
NSEM [N-Concept inanimate]													
NTYPE	<table border="0"> <tr> <td>[NSEM [COMMON count]]</td> <td></td> </tr> <tr> <td>NSYN common</td> <td></td> </tr> </table>	[NSEM [COMMON count]]		NSYN common									
[NSEM [COMMON count]]													
NSYN common													
1	CASE nom, GEND masc, N-FORM obl, NUM sg												
TNS-ASP	[MOOD indicative, PERF -, PROG -, TENSE pres]												
34	AUXTYPE tho, GEND masc, NUM sg, PASSIVE +, VTYP main												

Figure IV - 8: F-structure Analysis of a Sentence with Passive Verb.

2.1.5 Test File 5

Features covered in Test file 5 include sentence level adjuncts, noun phrase level adjuncts, and SUBJ, OBJ and COMP subcategorization. Adjuncts include postpositional phrases and adjective phrases. Total of 84 sentences are there in Test file 5. A random sentence analysis from Test file 5 with SUBJ, OBJ and NP level adjunct is shown in Figures IV-9 and IV-10.

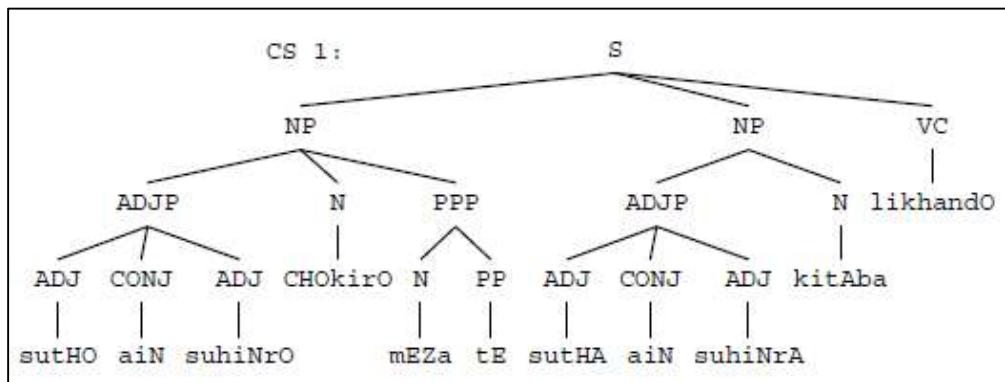


Figure IV - 9: C-structure of a sentence with Phrase level and Sentence Level Adjuncts.

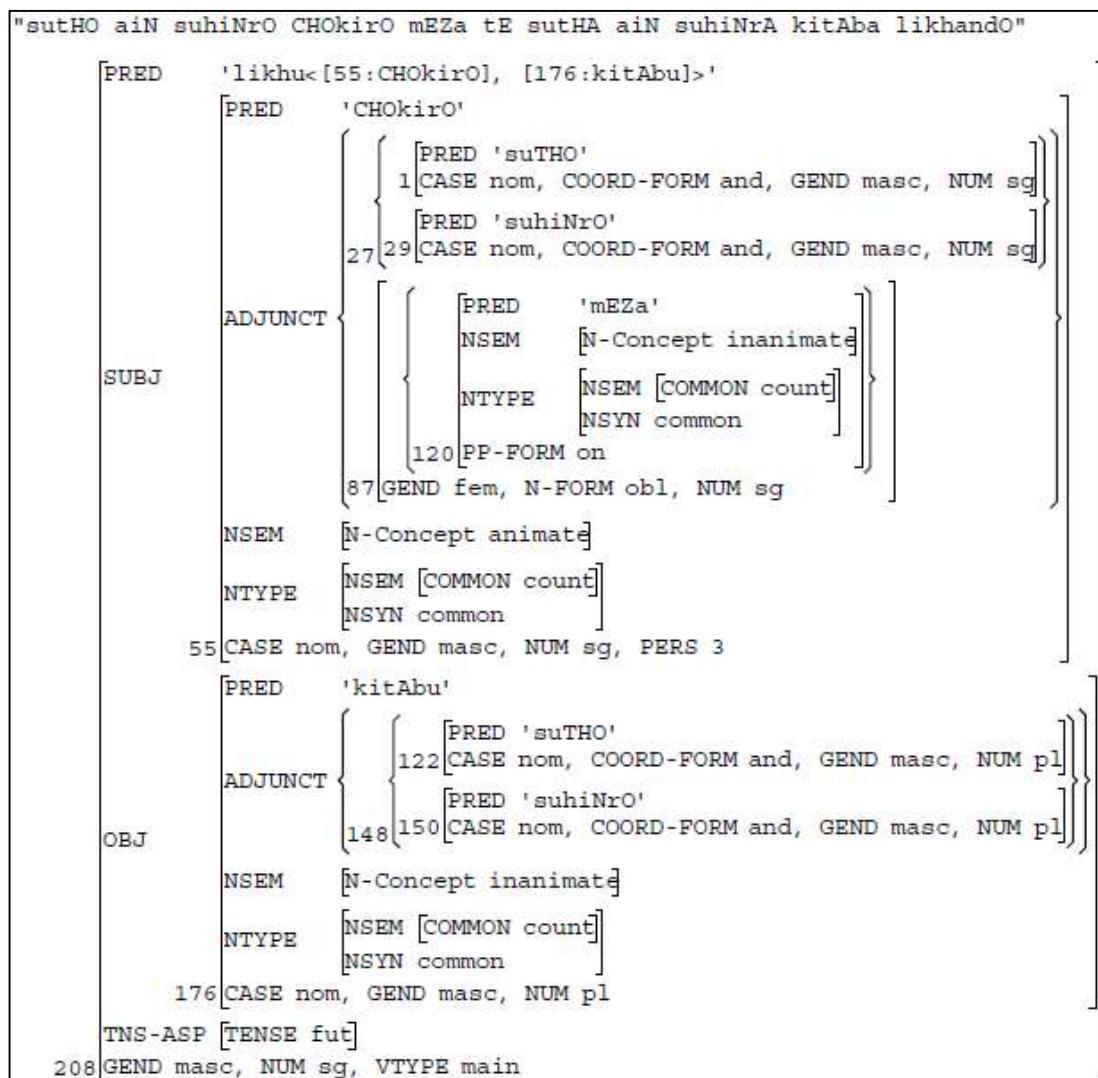


Figure IV - 10: F-structure Analysis of a Sentence with Phrase and Sentence Level Adjuncts.

2.1.6 Test File 6

Test file 6 covers SUBJ, OBJ, and XCOMP subcategorization along-with infinitives. Total 43 sentences are there in Test file 6. An example sentence analysis from Test file 6 with SUBJ, XCOMP subcategorization is shown in Figures IV-11 and IV-12.

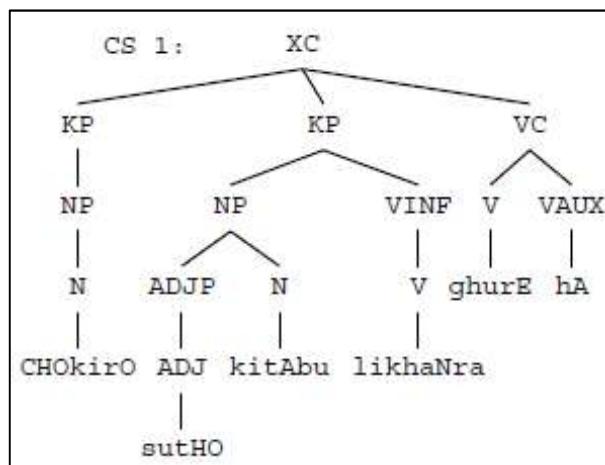


Figure IV - 11: C-structure of a sentence with SUBJ and XCOMP.

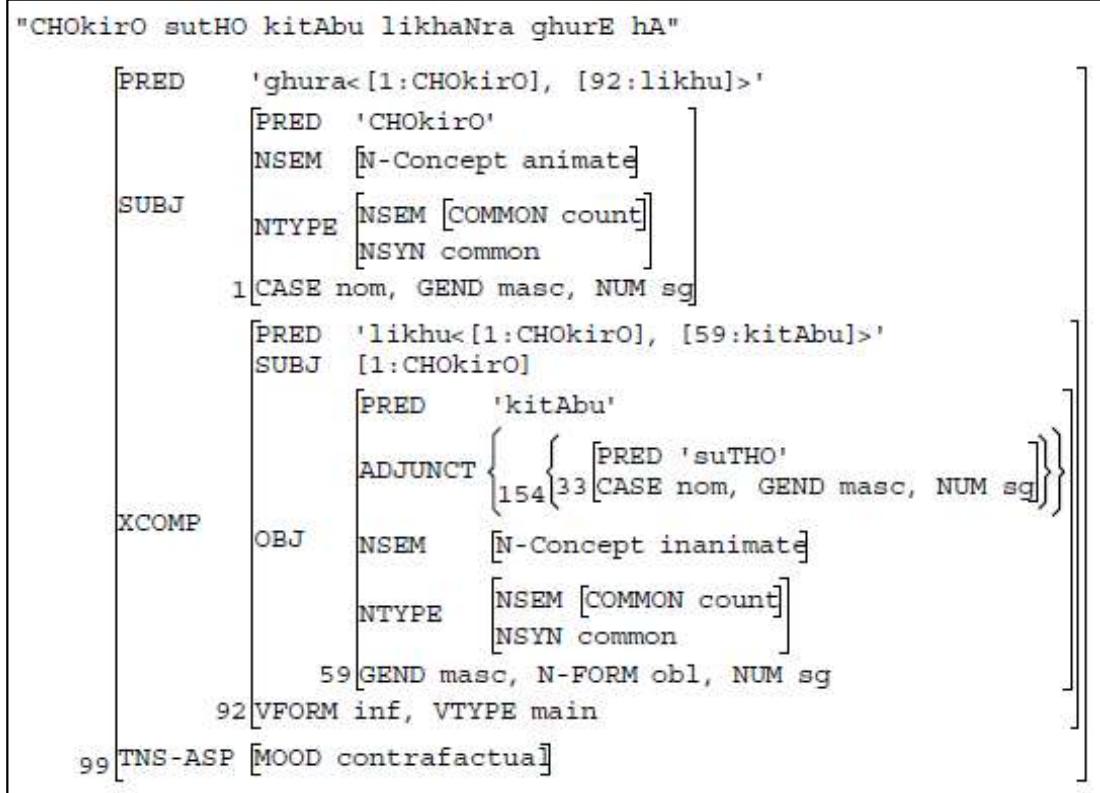


Figure IV - 12: F-structure Analysis of a Sentence with SUBJ and XCOMP Subcategorization.

2.1.7 Test File 7

Syntactic case is covered in Test file 7 sentences. A total of 122 sentences cover accusative, dative, locative, participant, genitive, and ablative cases. An example sentence analysis from Test file 7 with genitive case is shown in Figure IV-13 and Figure IV-14.

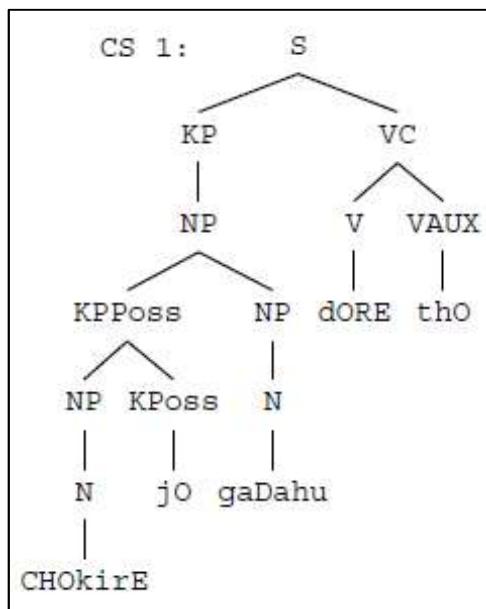


Figure IV - 13: C-structure of a Sentence with Syntactically Formed Genitive Case.

```

"CHOKirE jo gaDahu dORE tho"
PRED      'dORI<[35:gaDahu]>'
    PRED  'gaDahu'
        NSEM  [N-CONCEPT animate]
            NTYPE [NSEM [COMMON count]
                      NSYN common]
SUBJ          PRED  'CHOKiro'
            NSEM  [N-CONCEPT animate]
                SPEC   [NSEM [COMMON count]
                            NSYN common] GEND masc, N-FORM obl, NUM sg,
1 CASE gen, CGEND masc, CNUM sg, PP-FORM of
35 CASE nom GEND masc, N-FORM obl, NUM sg, PERS 3
TNS-ASP [MOOD indicative, PERF -, PROG -, TENSE pres, TENSE-FORM aorist]
69 AUXTYPE tho, GEND masc, NUM sg, VTYPE main

```

Figure IV - 14: F-structure Analysis of a Sentence with Syntactically Formed Genitive Case.

2.1.8 Test File 8

Test file 8 covers nested complement phrases with SUBJ, OBJ and COMP subcategorization in total 12 different sentences. A sample sentence analysis with nested complement phrase is shown in Figures IV-15 and IV-16.

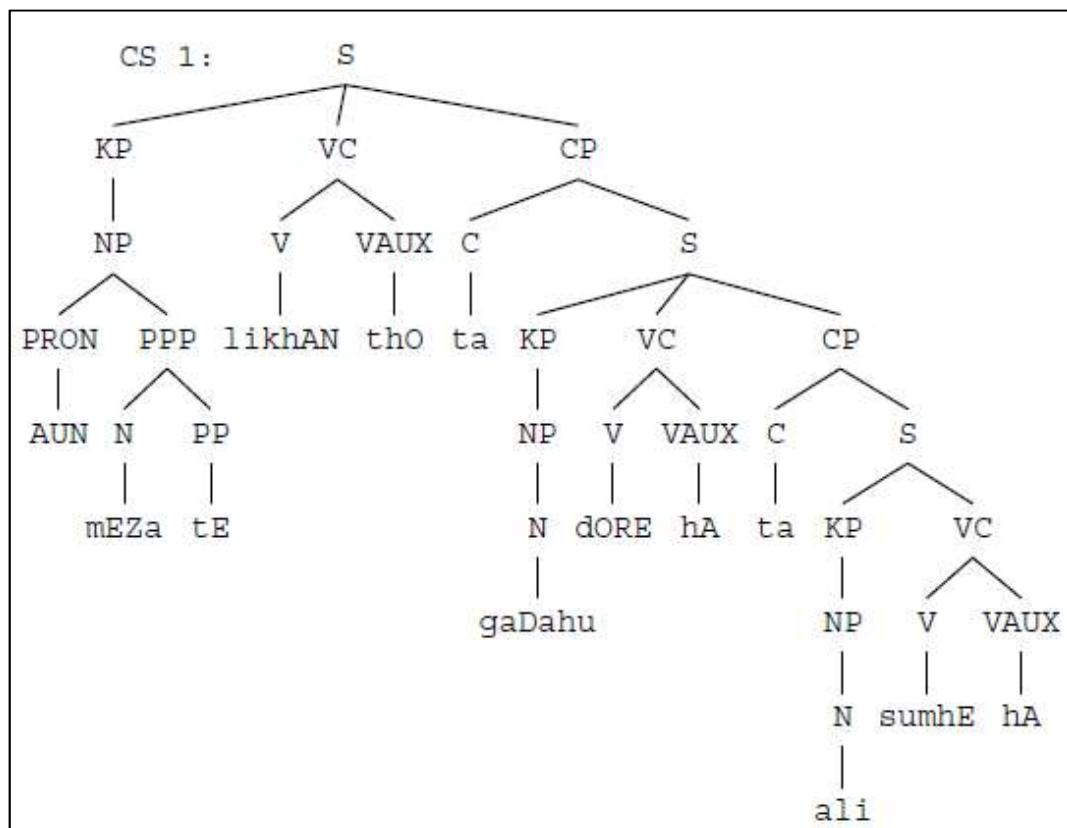


Figure IV - 15: C-structure of a Sentence with Nested Complement Phrase.

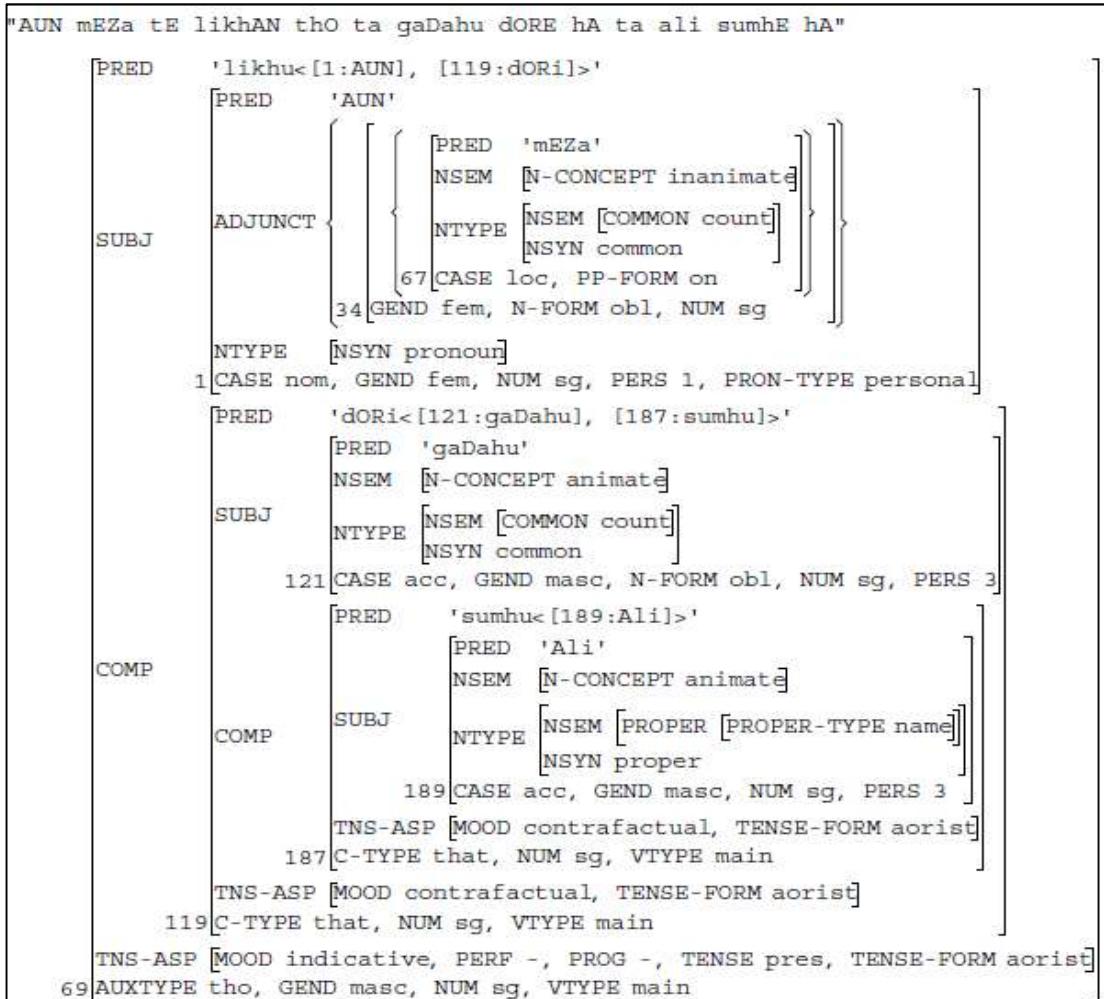


Figure IV - 16: F-structure Analysis of a Sentence with Nested Complement Phrase.

2.1.9 Test File 9

Subject pro-drop phenomenon with dummy pronoun subjects is covered in Test file 9 containing 18 different sentences. C-structure and f-structure analysis of a sentence with pro-dropped subject is shown in Figure IV-17 and Figure IV-18 respectively.

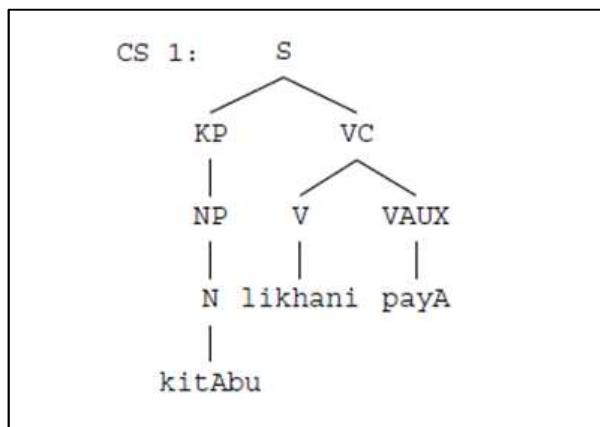


Figure IV - 17: C-structure of a Sentence with Pro-dropped Subject.

"kitAbu likhani payA"	
PRED	'likhu<[34-SUBJ:pro], [1:kitAbu]>'
SUBJ	[PRED 'pro' CASE nom, NUM pl, PERS 3]
OBJ	[PRED 'kitAbu' NSEM [N-CONCEPT inanimate] 1 104 NTYPE [NSEM [COMMON count]] 56 106 NSYN common 135 34 113 CASE acc, GEND masc, N-FORM obl, NUM sg]
125 137	TNS-ASP [MOOD indicative, PROG +, TENSE pres, TENSE-FORM aorist]
138	AUXTYPE payo, GEND masc, NUM pl, VTYPE main

Figure IV - 18: F-structure Analysis of a Sentence with Pro-dropped Subject.

2.1.10 Test File 10

Test file 10 covers pronominal suffixes and pro-drop caused by pronominal suffixes including subject and object pro-drops. An example object pro-dropped sentence analysis is shown in figure IV-19 and IV-20.

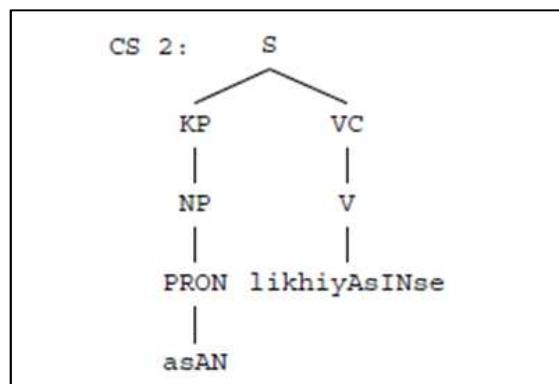


Figure IV - 19: C-structure of a Sentence with Object Pro-drop Caused by
Pronominal Suffixation.

```

"asAN likhiyAsINse"
PRED      'likhu<[1:asIN], [34-OBJ:pro]>'
          PRED  'asIN'
SUBJ      NTYPE [NSYN pronoun]  PERS 1, PRON-TYPE personal
          1CASE nom, GEND masc, N-FORM obl, NUM pl,
OBJ       PRED 'pro'
          CASE acc, GEND fem, NUM sg, PERS 3
TNS-ASP  [PERF +]
34NUM pl, PTCPL-TYPE past, VTYPE main

```

Figure IV - 20: F-structure Analysis of a Sentence with Object Pro-drop
Caused by Pronominal Suffixation.

2.2 Test Suite 2

Two Sindhi text books of class 1 are taken as test suite 2. Every text book is represented as a test file. Test file 1 contains 117 sentences taken from “siNdHI BArANrU kitAbu” (Sindhi Nursery Book). Test file 2 contains 141 sentences from “sindHI pahriyUN kitAbu” (Sindhi Book One). Total 258 sentences are considered for Test Suite 2. Different grammatical features covered by test file 1 include subject and object pro-drop, SUBJ, OBJ, COMP and PREDLINK sub categorization, ADJUNCT and XADJUNCTs. Test file 2 coverage include Pro Drop, SUBJ, OBJ, PREDLINK, Sentence ADJUNCTS, NP-ADJUNCTS, NP-COORDINATION, ADJ-COORDINATION, PASSIVES, XADJUNCTs, and XCOMP. Snapshots of text book pages used in test file 1 and test file 2 along-with transliterated text are shown in Figure IV-21 and Figure IV-22 respectively.

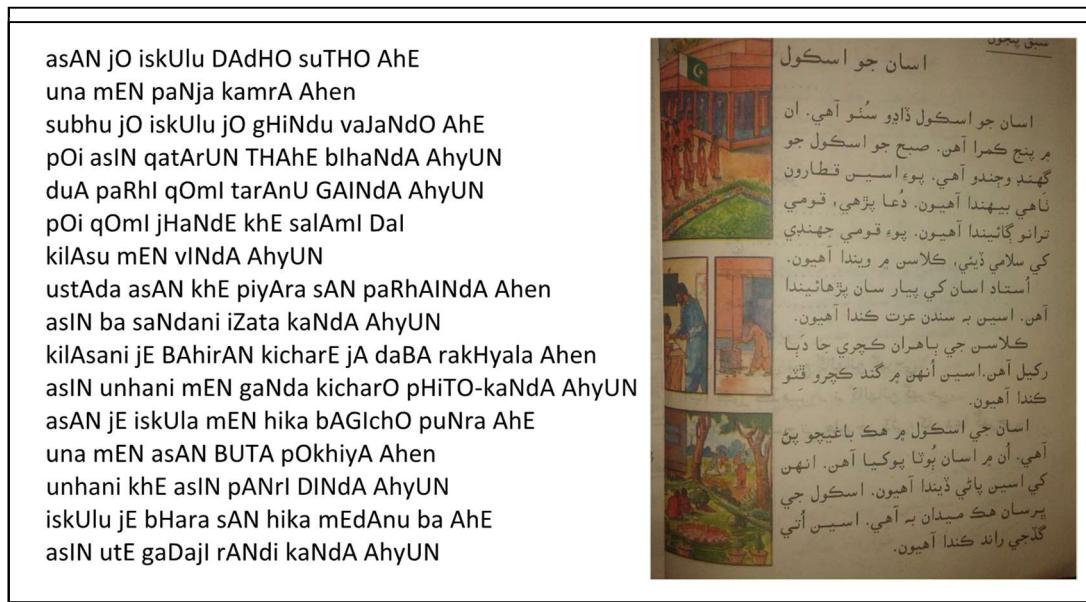


Figure IV - 22: Snapshot of a Page from Sindhi Book One and Transliterated Text Used in Test File 2 of Test Suite 2.

Transliterated Text Used in Test File 1 of Test Suite 2.

LFG analysis of a random sentence from Test file 1 of Test suite 2 is shown in Figure IV-23 and Figure IV-24.

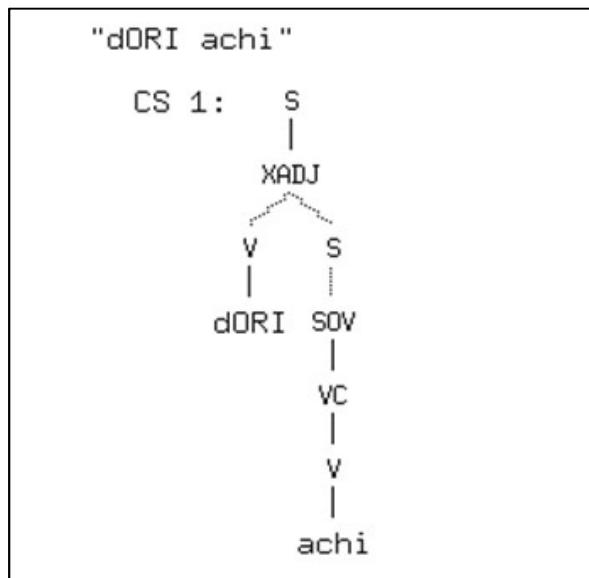


Figure IV - 23: C-structure of a Sentence from Test File 1 of Test Suit 2 with Pro-dropped Subject, and XADJUNCT.

"dORI achi"	
PRED	'dORi<[1-SUBJ:pro]>'
SUBJ	[PRED 'pro' NUM sg, PERS 2]
XADJUNCT	[PRED 'acha<[1-SUBJ:pro]>' SUBJ [1-SUBJ:pro] 34[NUM sg, V-Form imperative, VTYPe main]
1PTCPL-TYPE	conjunctive, VTYPe main

Figure IV - 24: F-structure Analysis of a Sentence from Test File 1 of Test Suit 2 with Pro-dropped Subject, and XADJUNCT.

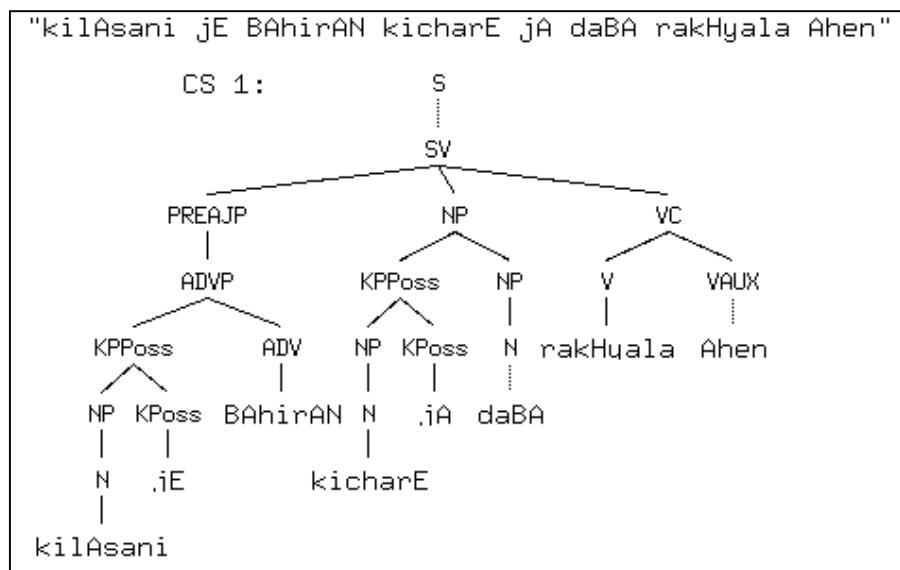


Figure IV - 25: C-structure of a Sentence from Test File 2 of Test Suit 2 with Sentence Level Predicative Adverb ADJUNCT.

"kilAsani jE BAhirAN kicharE ja daBA rakHyala Ahen"

```

PRED   'rakHu<[78:da80]>'          ]
|       PRED   'da80'                  ]
|       NSEM  [N-CONCEPT inanimate]    ]
|       NTYPE [NSEM [COMMON count]]    ]
|       NSYN  common                 ]
|       ]                                ]
SUBJ   [PRED   'kichar0'              ]
|       NSEM  [N-CONCEPT inanimate]    ]
|       NTYPE [NSEM [COMMON count]]    ]
|       NSYN  common                 ]
|       ]                                ]
SPEC   [PRED   'k'                    ]
|       NSEM  [N-CONCEPT inanimate]    ]
|       NTYPE [NSEM [COMMON count]]    ]
|       NSYN  common                 ]
|       ]                                ]
|       K-GEND masc, K-NUM sg, N-FORM obl
43CASE gen, GEND masc, K-FORM voc, NUM sg, PP-FORM of.
78CASE voc, GEND masc, NUM sg, PERS 2
|       ]                                ]
ADJUNCT [PRED   'BAhirAN<[1:kilAsu]>'      ]
|       PRED   'kilAsu'                ]
|       NSEM  [N-CONCEPT inanimate]    ]
|       NTYPE [NSEM [COMMON count]]    ]
|       NSYN  common                 ]
|       ]                                ]
OBJ    [PRED   'ba'                   ]
|       NSEM  [N-CONCEPT inanimate]    ]
|       NTYPE [NSEM [COMMON count]]    ]
|       NSYN  common                 ]
|       ]                                ]
|       K-GEND masc, K-NUM sg, N-FORM obl,
1CASE gen, GEND masc, K-FORM obl,           NUM pl, PP-FORM of.
36ATYPE space
|       ]                                ]
TNS-ASP [PERF +, TENSE-FORM aorist]
40GEND masc, NUM pl, PERS 3, PTCPL-TYPE past, VTYPE main

```

Figure IV - 26: F-structure Analysis of a Sentence from Test File 2 of Test Suit 2 with Sentence Level Predicative Adverb ADJUNCT.

3. RESULTS

As explained earlier that in 10 different test files of Test suite 1 total 617 sentences are used for parsing. As a result, total 605 sentences were parsed successfully. A bar chart showing results of test suite 1 is given in Figure IV-27. In two test files of Test suite 2 total 258 sentences were there and 249 were successfully parsed. Parsing results of individual files of test suite 2 are shown in bar chart of Figure IV-28. Figure IV-29 and Figure IV-30 show bar charts of individual test suite results and consolidated results respectively.

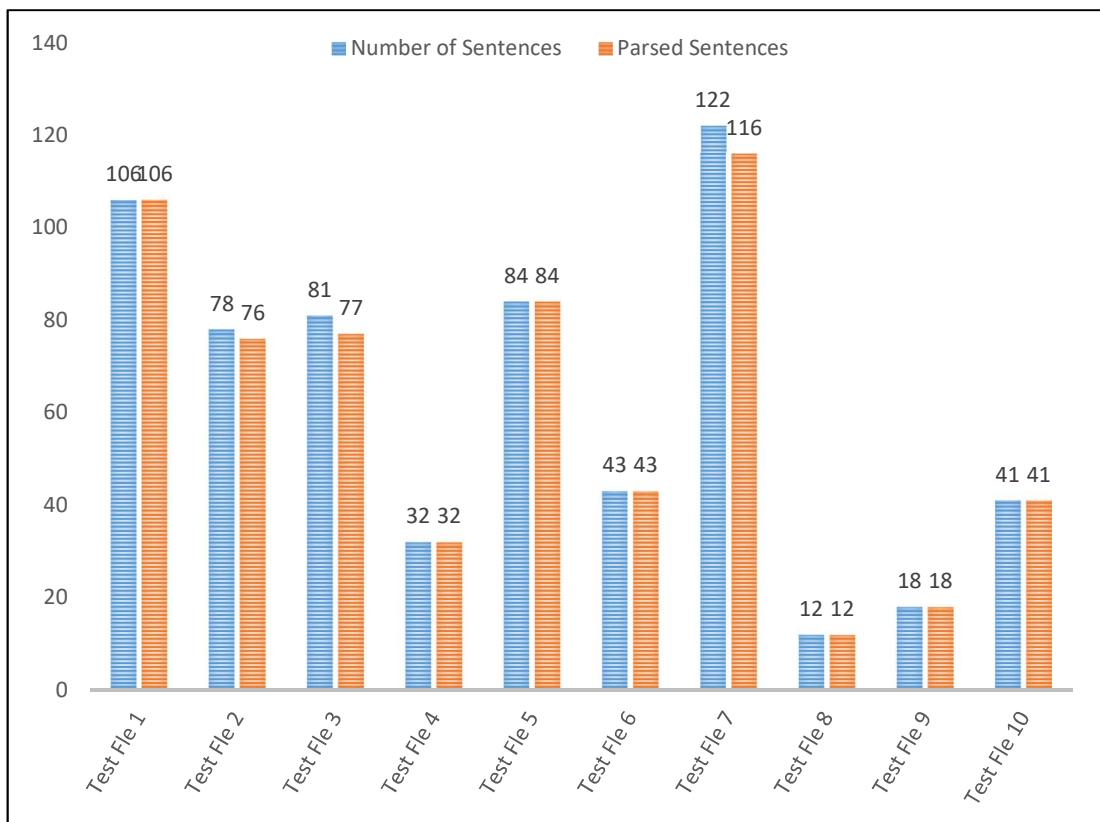


Figure IV - 28: Parsing Results of Individual Files of Test Suite 1.

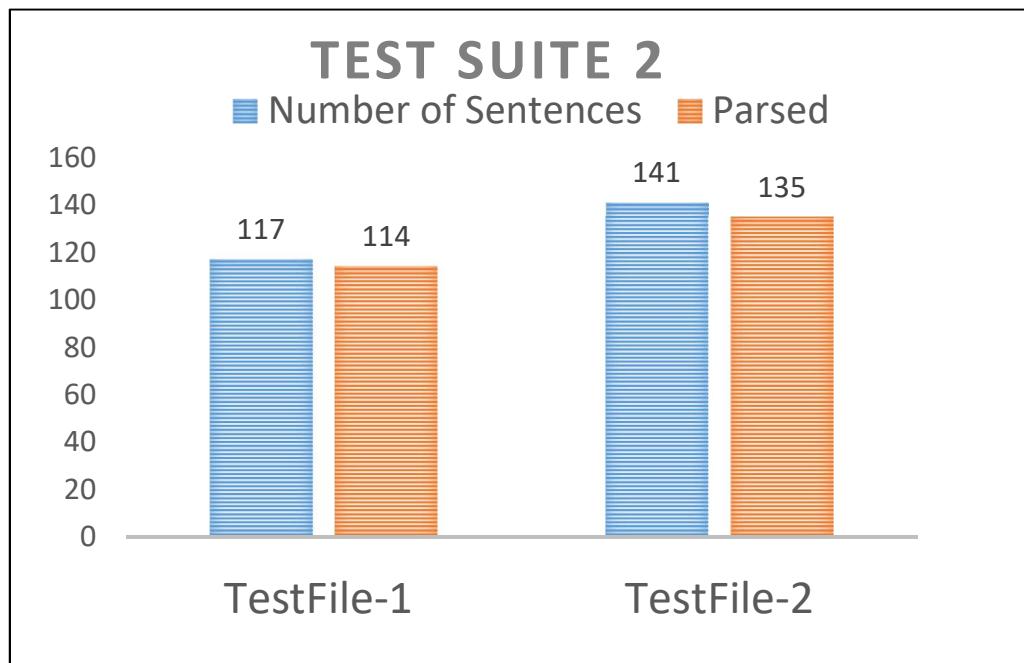


Figure IV - 27: Parsing Results of Individual Files of Test Suite 2.

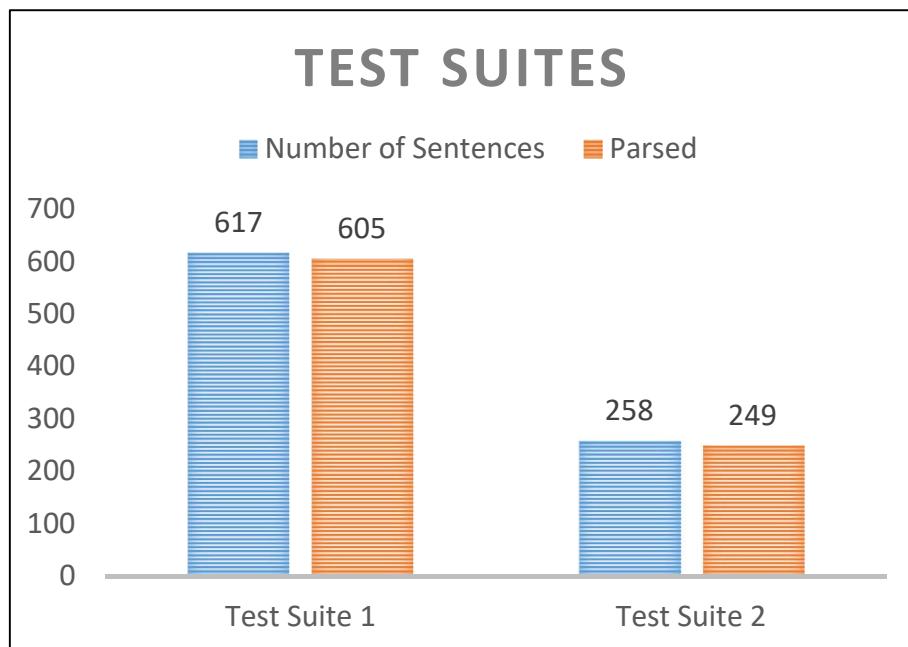


Figure IV - 29: Parsing Results of Individual Test Suites.

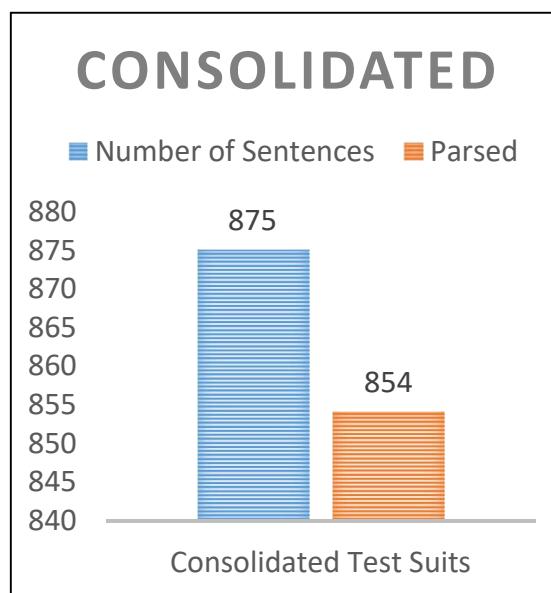


Figure IV - 30: Consolidated Parsing Results of Test Suite 1 and 2.

3.1 Unparsed Sentences

Sentences not parsed in Test Suite 1 & 2 were either bad sentences or having unhandled phenomenon. Different reasons of unparsed results include

NP coordination (i.e. sentences look fine but are not grammatical like (1) in following Table IV-5), individual use of demonstrative pronouns, post verbal modifications like intensifiers, complex PREDLINKs, and complex verbal constructions. Various patterns of unparsed sentences are given in Table IV-5 with explanations.

4. GRAMMAR COVERAGE

LFG rules coverage is shown in Table IV-6. Total 24 rules are implemented and are used to parse the sentences in test suites discussed above. Most of rules are completely used along-with their sub-rules / choices.

However, few rules are partially used as their sub-rules or choices are not used completely due to various reasons. Table IV-7 shows the list of completely used LFG rules along-with their description. Table VII-8 shows the list of partially used LFG rule along-with number of un-used choices, and reason of un-used choices. Reason behind most of the un-used rules is template invocation with all choices (P1, P2 and P3) of pro-drop templates.

Table IV - 5: Unparsed Sentences with Explanation.

S. No.	Sentence	Explanation
1.	tavhIN aiN uhE dORO payA	Eventually bad (agreement problem) due to coordination NP becomes in 3P and verb “dORO” is 2P.
2.	hO dORani	“hO” is demonstrative pronoun can’t be used standalone must be used with Noun or NP as specifier.
3.	puJi ta saHI	“ta” and “saHI” are post verbal modifiers. Post verbal modifiers are not handled.
4.	gaDahu KHaCHaru jEDO na AhE	Post nominal Pronouns / Adjectives in PREDLINK are not handled. “jEDO” can either be post nominal adjective or pronoun.
5.	plaNra IAi pANrl SAfa huaNra gHuriJ E	Infinitive Copula PREDLINK as XCOMP (SAfa huaNra) not handled
6.	mANrahU malAkHaRO DisaNra IAi vaNGaNra laGA	Complex Verbal Construction with XCOMP (DisaNra IAi vaNGaNra laGA) not handled Subject to further research
7.	janvara ba pehiNjE vatHANrani tE mOTI INdA Ahen	Complex verbal construction (mOTI INdA Ahen) left as open research question
8.	pARE jUN CHOKiryUN Sadaf jE gHaru achI gaDu thiYUN	Complex verbal construction, however can be parsed if broken into two sentences but left as open research question

Table IV - 6: Grammar Coverage.

Total Number of LFG Rules	24
Coverage by test Corpus	24
Partially Un-used Rules	6
Unused choices in 6 Partially Unused rules	87

Table IV - 7: Completely Used LFG Rules.

S. No.	Grammar Rule	Description
1.	S	Sentence Root for All sentences
2.	SV	Subject Verb Sentence Constructions
3.	PREAJP	Pre-Adjunct Phrase
4.	POSTAJP	Post-Adjunct Phrase
5.	NPRel	Noun Phrase with Relative Clause
6.	NP	Noun Phrase
7.	CompoundNP	Compound Noun Phrase
8.	NPDual	Dual Noun Phrase
9.	InfNP	Infinitive Noun Phrase
10.	ADJP	Adjective Phrase
11.	PPP	Postpositional Phrase
12.	ADVP	Adverbial Phrase
13.	KP	Case Phrase
14.	XCP	XCOMP Phrase
15.	KPPoss	Possessive Case Phrase
16.	CP	Complement Phrase
17.	KPPossinf	Possessive Infinitive Case Phrase
18.	VC	Verbal Complex

Table IV - 8: Partially Used LFG Rules.

S. No.	Rule	Description	Unused Choices	Reason
1.	SOV	Subject Object Verb Sentence Constructions	21	Optional Pro-drop choices are un used
2.	SOOV	Root for Subject, Object, Object 2, and Verb Constructions	5	Optional Pro-drop choices are un used
3.	SOBLV	Subject, Oblique and Verb Constructions	25	Optional Pro-drop choices are un used
4.	SOOOV	Subject, Object2, Oblique, Object Constructions	30	Choices unused due to optionality and Pro-drop
5.	CNP	Coordinated Noun Phrase	2	unused choices in MACRO
6.	NPPRONOUN	Pronoun Noun Phrase	4	choices. Few pronoun types were never used.

5. OTHER OUTCOMES OF RESEARCH AND DEVELOPMENT

While developing the morphology and LFG grammar of Sindhi a part of speech tagset and roman script for Sindhi is also developed. Tagset was designed while finalizing the morphological attributes of different parts of speech in Sindhi. The roman script which is used throughout the research study was also designed on almost same guidelines on which Urdu roman

script is used in ParGram project (Kamran, et al., 2010). Proposed tagset and roman script are shown in Appendix A and B respectively.

CHAPTER V

DISCUSSION

This research study focuses the development of Sindhi morphology and grammar within FSM and LFG frameworks respectively. Initially a comprehensive survey of Sindhi linguistic patterns is carried out through an extensive literature review of existing Sindhi grammar material. It was discovered that many conventional grammatical constructions defined by Sindhi grammarians need to be revised with modern linguistic theory perspective (LFG in this case). In Chapter IV “Methods and Tools”, Sindhi word classes and their sub-types are defined along-with their characteristics. While defining sub-types and their characteristics various decisions are made based on different observations. For example, Sindhi grammarians define compound postpositions (see Table III-3) which are actually embedded simple postpositions. In the same way six different types of conjunctions are defined as given in Section 1.1.3; however, all these conjunctions fall into two categories subordinate and coordinate conjunctions as shown in Table III-4. Different examples of Sindhi compound verbs and participles are analyzed and their patterns are defined in Table III-8 and III-9 respectively.

Five to eight different cases are reported by Sindhi grammarians; again, the analysis and survey revealed that there are ten different cases of noun in Sindhi. Simple postpositional case is actually divided into accusative, dative, locative, instrumental, participant, agentive, ablative and genitive cases (See sections 2.3 and 3.2 of Chapter III). Instrumental and participant cases are marked by same marker ‘sAN’, however, if noun is animate noun then the case

is defined as participant case otherwise it is instrumental. In the same way, ablative case marking can either be morphemeal or syntactic.

Native adjectives are declined like nouns with number, gender, and case marking (See Table III-16); however, adopted adjectives neither inflect nor marked by number, gender and case.

Pronouns are usually declined for number, rarely for gender and case. However, in case of wh-pronouns there are some exceptions for example wh-pronoun “kehRO” (which) is marked like common masculine nouns. This can be seen in Table III-20. Same is the case with indefinite pronouns given in Table III-22.

Postpositions have interesting inflections, for example, postpositions which are used as locative case markers are marked by ablative case and postpositions which are used as genitive case markers are marked by number, gender and case like nouns as shown in Table III-23. Manner adverbs inflect like nouns and adjectives and hold syntactic agreement properties.

Morphologically verb has very complex constructions, verbs in Sindhi are marked by number, gender, tense, aspect and mood along-with imperative, infinitive, participle, transitive and causative formations. Imperatives in Sindhi can have different formations which include default second person imperative, polite imperative, and desiderative. Morphological rules are defined for these imperative formations (excluding default second person imperative which is considered root) as discussed in section 2.9 of Chapter III. In the same way infinitive and participle formations are defined by different morphological rules. An interesting participle type in Sindhi is conjunctive participle which is derived from verb, acts as a conjunction and

marks open adjuncts (XADJUNCT) in Sindhi sentences. This can be seen in examples (9a – 9c) and Section 4.6.6. An interesting causativity pattern discussed in conventional Sindhi grammar is with “-rA-rAi” reduplications as discussed in Section 2.12; this reduplication increases the transitivity of verbs theoretically to any number of transitions. Corpus analysis however does not provide any evidence of such constructions. Tense aspect and mood are less studied parts of Sindhi grammar specially one cannot find material on aspect and mood patterns in Sindhi. Sindhi verbs have special form called aorist or old-present form without any aspect or tense and is marked by person and number as shown in Table III-36. This form provides basis for tense and aspect formations. Present tense is formed by aorist followed by present tense auxiliary which is further conjugated by number and gender as shown in Table III-37. Past tense formation have different patterns for transitive and intransitive verbs as shown in Tables III-40 and III-41. Future tense formations also have different patterns discussed in 2.13.4.

Aspect can either be perfective or imperfective and is marked syntactically. Imperfective aspect is further divided into habitual and continuous types. First pattern of perfective aspect found in Sindhi is with perfective-past-participle followed by auxiliaries in different tenses. Examples of this pattern are shown in Table III-47 and example sentences are shown in examples (11a – 11d). Second pattern of perfective aspect found is with conjunctive-participle followed by completion marker “vartO” as shown in Table III-48. Examples of this pattern are shown in sentences (12a – 12c). Two imperfective continuous aspect patterns were found where different continuity markers (payO and rahiyO) are used to mark continuous aspect. “payO”

follows aorist base in present tense, perfective participle in past tense and future form of verb in future tense as shown in Table III-49 and examples (13a – 13d). Imperfective aspect marked by “rahiyO” marker has common pattern i.e. conjunctive-participle+“rahiyO” followed by tensed auxiliary in different tenses as shown in Table III-50. Imperfective habitual aspect patterns for different tenses can be seen in Table III-51; present participle form of verb plays central role in habitual formation. Habitual aspect in different tenses differs only by tensed auxiliary which follows common present participle pattern. Interesting patterns of combined aspects are also discovered where continuous and habitual aspects are combined and resulting patterns have properties of both aspects. Two different pattern sets of aspect combinations in different tenses are discovered and shown in Table III-52 and III-53 respectively. Along-with four mood patterns defined by J. S. Cole (2001) six other mood patterns are observed and total 10 different moods are identified as discussed in Section 2.14. Imperative is marked morphologically by second person imperative form. Subjunctive is marked morphologically as well as syntactically as shown in Table III-54. Declarative, permissive, prohibitive, capacitive, suggestive and compulsive moods are marked syntactically by using different patterns discussed in different subsections of Section 2.14.

Sindhi pronominal suffixes are analyzed and different pronominal suffixation patterns are defined. Pronominal suffixes can appear with nouns, verbs, auxiliaries and postpositions as shown in Tables III-(58 – 62). Double pronominal suffixation may appear with verbs, these patterns can either have 1st Person – 2nd Person and 1st Person – 3rd Person sequences as shown in Table IV-63 or 3rd Person – 1st Person, 3rd Person – 2nd Person and 3rd

Person – 3rd Person sequences as shown in Table III-64. Nouns and adverbs can also have postpositional suffixes. When used with nouns these suffixes either reflect ablative or participant case of nouns as shown in examples (32a – 32b) the ablative case of nouns or adverbs. Postpositional suffixes with adverbs of place reflect the ablative case of place adverbs (see Table III-65).

Noun phrase constructions in Sindhi may include simple or complex constructions. Simple constructions include either single noun or pronoun; however, complex NP constructions may include pronoun-noun, adjective-noun, pronoun-adjective-noun, postpositional phrases, relative clauses and their coordination. NP definition rules and example constructions are given in Section 3.1. Extensive discussion on different patterns of NP and their example sentences can be seen in discussion on examples (33a – 39).

Case marking on different NP constructions is defined through special case phrase (KP) in section 3.2.12. Apart from nominative (which is default morphological case in Sindhi) and vocative cases all other cases are marked syntactically through oblique morphological form followed by a case marker. It may also be noted that Sindhi do not have ergative case marking and ergativity is reflected by oblique nominal form as shown in examples (42a – 42b). Differential object marking phenomenon is also observed in Sindhi where accusative case is either marked by case marker “kHE” or remain unmarked in nominative form with subject reflecting the ergativity in oblique form; this can be seen in examples (44a – 44c). Genitive case is marked by genitive case marker “jO” following the oblique form. Genitive is different than other cases as it contains possessor and possessed nouns and holds syntactic agreement between case marker and possessed noun as shown in examples (53a – 53c).

With first and second person pronouns genitive is also marked morphologically (see examples 54a and 54b); however, syntactic agreement holds and pronoun morphological form must agree in number, gender and case with possessed noun.

Different verbal subcategorization patterns for Sindhi in LFG theory are defined for various grammatical functions including subject, object, secondary object, oblique, complement, and open complement. Thematic hierarchy defined by (Bersnan and Dalrymple, 2001) is used for identification of different grammatical function. However, argument identification problem discussed in example (58) remains unsolved and need further investigation. As discussed earlier that interesting pattern of open adjunct (XADJUNCT) found in Sindhi is marked by conjunctive participle; this can be seen in example (65).

Results and evaluation of implemented morphology and syntax are given in Chapter IV. As shown in Table IV-1 stem forms coverage implemented includes 418 noun stems with 2086 surface forms (word forms which are generated from these stems). Verb is morphologically rich word class and one verb can have up to 75 different surface forms as shown in Table III-75 in Chapter III. This richness is also reflected in Table IV-1 where 136 verb stems can generate 7339 surface forms. In the same way stem and surface forms of all word classes implemented in LEXC are listed in Table IV-1. A total of 862 stems are implemented with 10327 surface forms. Various categories identified for different morphological classes as per their morphological patters are listed in Table IV-2. Different word classes are subcategorized according to inflectional categories as these categories have different morphological rules. Table IV-2 also lists average inflections per stem for each word class.

Total 31 noun categories are defined with 4.99 average inflections per stem. Interestingly adjectives have more average inflections per stem as compared to nouns (5.64 vs 4.99); this is due to double inflections when adjectives are further inflected on degree. For example, the adjective “tHOr-O” will have all inflections (number, gender and case) like a masculine noun with “O” ending plus other inflections which are caused by change in degree value of adverb like “tHOr-aR-O” in this case. Now adjective form “tHOr-aR-O” will also go through number, gender and case inflections; so, if a noun with “O” ending has 12 inflections then adjective with “O” ending may have up to 24 or more inflections. A total of 86 inflectional categories are defined with 11.80 average inflections.

Evaluation of the developed finite state morphology is done in terms of precision (fraction of words for which correct analysis is given i.e. number of correct findings divided by total number of findings), recall (fraction of expected correct analyses i.e. number of correct findings divided by number of expected findings or gold standard), coverage (number of morphological forms for which at least one analysis is returned), and average ambiguity (average number of analyses returned for every word). Precision, recall, f-measure (f1), coverage and average ambiguity results are shown in Table IV-3. It can be observed that ambiguity is affecting the precision percentage; for example, pronouns have highest ambiguity and lowest precision and adverbs have lower ambiguity with higher precision. A comparative analysis with Apertium is given in Table IV-4 shows that Apertium implementation is slightly better in terms of f1-measure and Xerox is better in terms of ambiguity.

Evaluation of developed grammar is done in terms of parsing and grammar coverage against two different sets of sentences. First set contains ten different files with sentences covering various linguistic features. Second set contains sentences from two books of class one. In Chapter IV section 2.1 descriptions of different test files along-with linguistic features coverage are given. Test file 1 covers PREDLINK and copula verbs with different number and gender paradigms. Pronominal subjects are also covered. Figure IV-2 shows analysis of a sentence of test file 1. Copula verb “AhE” construction with noun PREDLINK can be observed in this analysis. F-structure analysis shown in Figure IV-4 is another example of PREDLINK with coordinated adjective construction. Coordination of two adjectives “suhiNrO” (beautiful) and “suTHO” (good) can be observed here it can be seen that after coordination number attribute remains singular. Test file 3 covers grammatical agreement, personal pronouns, oblique forms and complex NP constructions. F-structure analysis of a sentence from test file 3 is given in Figure IV-6. It can be seen that subject is a complex noun phrase construction with coordination of four different nouns. Coordinated NP number becomes plural. In TNS-ASP attributes it can be observed that this sentence is an example of progressive (continuous) aspect in present tense. Passive forms of verbs are covered in test file 4. Figure IV-8 shows an example sentence from test file 4 with passive form of verb “likHu” in present tense and indicative mood. Phrase and sentence level adjuncts are covered in test file 5. F-structure analysis of Figure IV-10 shows complex NP construction with coordinated adjective phrase and postpositional phrase marked as adjuncts of subject noun “CHOkirO”. In the same way object NP also contains coordinated adjective phrase as an adjunct. Sentences with

open complement (XCOMP) constructions are given in test file 6. XCOMP is marked by infinitives. Analysis of a sentence from test file 6 is shown in Figure IV-12. The sentence contains two main verbs; infinitive form “likHaNra” and “gHurE”. Predicate stem “gHura” of verb “gHurE” has two arguments, a subject “CHOkirO” and an XCOMP verb “likHu” which is further subcategorized for subject and object. Subject of XCOMP is referred from main sentence subject so is the case here and subject of “likHu” is “CHOkirO” which is referred from main sentence predicate “gHura”. Object of XCOMP is book with an adjective adjunct. The auxiliary “hA” marks the contrafactual mood as shown in TNS-ASP attributes. Sentences with syntactic case formation patterns are given in test file 7. Example analysis given in Figure IV-13 and IV-14 is of genitive case as genitive is most complex case formation in Sindhi. Analysis shows that verb “dORi” has only one argument “gaDahu” in nominative case. The case phrase “CHOkirE jO” is treated as a specifier of subject “gaDahu” in genitive case. Closed complement (COMP) sentences are given in test file 8. Sentence analysis of Figure IV-16 shows that there are nested closed complements in the sentence. First of all, main verb “likhu” (write) is subcategorized for first person pronoun subject, and complement verb “dOri” (run). Next, verb complement “dORi” is subcategorized for subject “gaDahu” (donkey) and another complement “sumhu” (sleep). Then the intransitive verb complement “sumhu” is subcategorized for subject “Ali”. It may be noted that closed complements always have their own subjects. Sentences with pro-drop phenomenon are given in test files 9 and 10. Test file 10 covers pronominal suffixation which causes pro-drop. Subject pro-drop can be observed in Figure IV-18 where subject is reconstructed as a third person nominative, plural

pronoun. Object pro-drop is observed in Figure IV-20 analysis where pronominal suffix caused object pro-drop which is reconstructed as a singular, third person pronoun in accusative case during syntax analysis. Test suit 1 sentences were handcrafted sentences to evaluate the developed grammar on different grammatical features. Test suit 2 contained sentences from two books of class one. Figure IV-24 shows analysis of a random sentence of test file 1 of test suite 2. This sentence is an example of XADJUNCT with pro-dropped 2nd person singular subject. LFG analysis of another example sentence from Test file 2 of Test suite 2 is shown in Figure IV-25 and Figure IV-26. Sentence level adjunct with predicative adverb can be seen in this sentence. Evaluation in terms of parsing of sentences can be observed in graphs of Figure IV-27 to Figure IV-30. A total 875 sentences were used for parsing and 854 were parsed successfully; 21 sentences which were not parsed either had unhandled phenomenon or were bad sentences. Table IV-5 lists the reasons along-with unparsed sentence examples. Evaluation of in terms of coverage is shown in Table IV-6. Table shows total number of rules, number of rules used by test corpus, partially un used rules and number of choices which are not used in partially un-used rules. Partially used LFG rules their unused choices and reasons are listed Table IV-8.

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

This thesis discusses the development of Sindhi LFG by identifying the morphological and syntactic patterns of Sindhi. First a comprehensive survey of Sindhi word classes and their morphological patterns is carried out which includes literature review of existing literature of conventional Sindhi grammar and corpus analysis. Morphological patterns of different POS classes are identified. These patterns are then modeled in finite state transducers by using XFST LEXC and as a result a morphological lexicon for Sindhi is developed. A comprehensive survey of Sindhi syntax is also carried out and various nominal and verbal elements of Sindhi are studied with LFG perspective. An LFG grammar for Sindhi is developed by incorporating the morphological lexicon of Sindhi. The developed morphological lexicon and grammar are then tested and evaluated against hand crafted sentences and real time corpus of “Sindhi Class One” books.

Sindhi morphological constructions include derivational and inflectional morphology. Morphologically nouns are marked by number (Singular & Plural) gender (Masc, Fem) and case. Conventional grammarians of Sindhi discuss five to eight different cases in Sindhi. However, after having the morphology and syntax survey ten different cases of nouns are identified namely nominative, accusative, dative, participant, instrumental, locative, ablative, agentive, genitive and vocative. Nominative is default case of nouns without any inflection. Accusative, dative, participant, instrumental, locative, ablative and agentive cases are marked syntactically by different case markers

following the oblique morphological form of noun. Possessive or genitive case is also formed syntactically by oblique noun form followed by possessive case markers. However; genitive case is different as it is marked by special case markers which are further marked by number and gender; syntactic agreement between these markers and possessed noun holds. Proper nouns are not generally inflected but there are exceptions of proper noun inflections in Sindhi and the inflection pattern is same like common nouns. When number, gender and case morphology is combined it generates up to 12 different morphological forms of a noun. A total of 31 different categories of nouns are identified and modeled in LEXC along-with their morphological paradigms. Native Sindhi adjectives like "suhiNrO" are declined like nouns but borrowed adjectives like "KHUbSUrat" are not declined. Adjectives can also have cardinal and multiplier inflections. Pronouns in Sindhi are declined for number and gender and are marked by default nominative case, oblique form and genitive case (genitive in pronouns can also have morphological as well as syntactic constructions). Most of the postpositions in Sindhi remain un-inflected, however, few are marked by number and gender and sometimes case (ablative case). Possessive post positions are marked by number, gender, nominative case, and oblique form and play important role in syntactic case marking of genitive case. Adverbs are generally not inflected neither they hold syntactic agreement but this is not the case with manner adverbs in Sindhi (i.e. adjectives when used as adverbs). Manner adverbs not only inflect but also hold syntactic agreement properties. Place adverbs can also have morphological ablative case. Conjunctions and interjections are neither inflected nor marked by any case. Verbs in Sindhi are morphologically complex

word class and marked by number, gender, aspect, tense and mood. Aspect and mood are less investigated parts of Sindhi grammar and are extensively explored in this study and tense, aspect and mood patterns are identified and modeled in morphology and syntax implementation. Verb root is always in second person imperative form which is further inflected to have more morphological imperative forms (polite, desiderative and requestive). Imperatives are also marked syntactically by using intensifiers or negations. Infinitives are derived by inflecting verb roots; inflections depend on last diacritic or vowel of verb root. Five different categories of morphological patterns of infinitive formation are identified and implemented. Participles are also derived from verb roots and there are five different participles in Sindhi namely present-participle, past-participle, future-participle, verbal-noun, and conjunctive participle. Present, future and past participles may have number and gender inflections as well. Intransitive verbs are inflected to form transitive verbs and transitive verbs can also inflect to form causative formations. Tense aspect and mood are marked morphologically as well as syntactically. Present tense formations are based on special form of verb called old-present or aorist by some grammarians. Present tense is marked by present tense auxiliaries following the aorist form of verb. Past tense is marked by past participle form of verb followed by past tense auxiliary/ies. Future tense is formed by present participle base which is further marked morphologically by number, gender and person. Aspect in Sindhi can either be perfective or imperfective, imperfective is further divided into continuous and habitual aspects. Perfective aspect can either be morphological or syntactic. Morphologically perfective aspect is marked by past-participle or perfective participle form of verb. Syntactically

perfective aspect is marked by conjunctive participle form of verb followed by perfective past participle auxiliary followed by tense auxiliary. Perfective aspect can be formed in present, past and future tenses and aspect of standalone aorist form of verb remains unspecified. Imperfective continuous aspect is formed syntactically; where base-forms of present, past and future tense (aorist, perfective-past-participle, and present participle respectively) are followed by continuity marker auxiliary. Alternate patterns also exist where conjunctive-participle and past participle of continuity marker auxiliary are followed by tensed auxiliary. Imperfective habitual formation pattern includes present participle and tensed auxiliary combination which may also be marked by gender, number and person. Patterns of aspect combinations are also found; for example, imperfective continuous and imperfective habitual aspects can be combined and combination will have continuity and habitual markers along-with tense auxiliaries. Another continuous-habitual formation pattern includes present-participle of main verb, present participle of continuity marker and tense auxiliary. Mood in Sindhi can be expressed by morphological form of verb or syntactically using auxiliaries. Subjunctive, presumptive, imperative, declarative, permissive, prohibitive, capacitive, compulsive, and suggestive mood patterns are identified. Subjunctive is further divided into subjunctive imperfective, subjunctive perfective and subjunctive conditional moods. Presumptive can also have habitual and perfective constructions. Imperative mood is expressed via imperative morphological form of verb. Different imperative forms are either imperative inflections or followed/preceded by intensity markers. Declarative mood is marked by copula verbs. Permissive mood is formed by infinitive form of verb followed by any inflected form of

permissive auxiliary. Prohibitive is formed by either preceding permissive pattern by negation or following the permissive by postposition and prohibitive verb. Capacitive mood pattern includes conjunctive participle followed by different forms of modal verb which express the capability of doing something. Suggestive and compulsive are also marked by modal verbs following the infinitive and future participle forms respectively. Implementation results show that a verb can have up to 75 different morphological forms.

Sindhi is one of few Indo-Aryan languages with pronominal suffixes which may appear on nouns to mark genitive possessor pronoun, on postpositions to mark pronouns with different postpositional cases, on verbs (including auxiliaries and modals to mark subjects, objects, and secondary objects, on copula verbs to mark subjects and sometimes on adverbs of place to mark the pronouns modified by adverbs. These pronominal suffixes may be considered argument replacing morphemes as they may cause pro drop in Sindhi.

Noun phrase constructions in Sindhi can be simple or complex which may include pronouns, adjective phrases, postpositional phrases, participle phrases, and relative clauses. Adjectives modifying nouns appear before nouns in a noun phrase. Adjectives can be coordinated and appear as an adjective phrase before nouns. Noun phrases are marked by case. Apart from default nominative case and vocative case other cases are marked syntactically with different case markers. Accusative and dative cases are marked by postposition “kHE” following the noun phrase in oblique form. Participant case is marked by “sAN” marker following the animate noun in oblique form, instrumental is also marked by “sAN” following the inanimate

noun. Ablative can either be marked by ablative case marker “kHAN” or morphological inflection of noun. “kHAN” is also used as agentive case marker. Case marking ambiguity is resolved at verbal subcategorization level. Case marking of noun phrase is implemented as a special case phrase (KP) which defines different formal constructions of noun phrase with different cases. Genitive case is different than other cases as it required agreement with possessed noun and genitive marker itself is marked by number and gender inflections. A special KPPoss (possessive case phrase) construction is used to define genitive case marking of noun phrase. Sindhi do not have ergative case marker and ergativity is reflected by oblique case of nouns. Sentences with ergative subjects can have un-marked objects without any case marker where differential object marking phenomenon is observed. In differential object marking direct object is either marked by accusative case marker or remain un-marked (in default nominative form).

While analyzing Sindhi grammar with LFG perspective verbal subcategorization frames are defined. Various grammatical functions of LFG theory are mapped to Sindhi clause elements. Subjects, objects, secondary objects, obliques, complements, open complements, adjuncts, open adjuncts and predlink constructions are identified and modeled. Bersanan & Dalrymple thematic hierarchy is used to identify different functions in a clause. Verbal arguments (subjects or objects) can be suppressed due to passivization or dropped due to pronominal suffixation.

Morphology is implemented in LEXC and LFG grammar is implement in XLE. Chapters III discusses implementation details of morphology and syntax. Morphological lexicon developed in LEXC includes 862 stem forms and 10327

inflectional forms. This lexicon is integrated with LFG grammar implemented in XLE which covers the basic constructions of Sindhi sentence.

LFG analysis results pose many new research questions in Sindhi linguistic patterns including complex verbal constructions, post verbal modifications, complex noun phrase constructions which pose case number and gender marking challenges when different elements of a noun phrase are coordinated. The developed model covers the basic constructions of XCOMP and XADJUNCT, however these constructions need further research and development to cope with more complex patterns. Verbal complex may include complex predicates and need to be analyzed further. Verbal subcategorization challenges include proper marking of different functional elements of clause including subjects, objects, secondary objects and oblique constructions as standard tests need to be designed for correct identification of these functional elements in Sindhi. Differential case marking phenomenon in Sindhi needs further investigation to identify and analyze differential subject and object marking patterns.

Developed finite state lexicon has applications in various domains of NLP and related areas. Morphological analysis is the first step in making natural language understandable by machines. Developed Sindhi lexicon can be incorporated in different applications like Sindhi spell checkers, text to speech systems, part of speech taggers, knowledge discovery applications, and deep linguistic analyzers.

Results of deep linguistic analysis of Sindhi sentences in LFG provide basis for Sindhi language understanding by machines. LFG analysis results

provide data which is linguistically expressive and plays important role in language understanding. These results are based on linguistic knowledge and generated results capture this knowledge at different levels including morphology, syntax and semantics. LFG analysis of Sindhi can be used to extract structures from collections of text. These linguistically rich structures can be given input to machine learning algorithms and this synthesis of deep linguistic analysis and machine learning can be used for more accurate feature extractions. Predicate argument structures generated by LFG can be used to extract semantic triples which are fundamental building blocks of knowledge representation in machine readable format. Language tutoring systems can use results of LFG analysis for language learners. Deep linguistic analysis of Sindhi along-with machine learning algorithms for better understanding and processing of Sindhi language will be a good research area to work on. Use of semantic triples generated by predicate argument structures has applications in semantic web, knowledge extraction and information processing. The other arguments (obliques, adjuncts, complements etc.) which are identified in verbal subcategorization via LFG analysis help in understanding relationships between different entities of text collections. Future research on developed grammar may also include work on incorporating optimality theory, and rewriting the grammar on ParGram guidelines.

REFERENCES

- AGERRI, R., ARTOLA, X., BELOKI, Z., RIGAU, G., & SOROA, A. (2015). Big data for natural language processing: a streaming approach. *Knowledge-Based Systems*, 79, 36-42.
- AHMED, T. (2007). Ablative, Sociative, and Instrumental Case Markers. In Proceedings of Conference of Language and Technology, University of Peshawar Bara Gali Campus.
- ALEXIADOU, ARTEMIS. (1997) Adverb placement. A case study in antisymmetric syntax. Amsterdam: John Benjamins.
- ALLANA. G. A. (2010) *Sindhi Boli jo Tashrihi Grammar “Descriptive Grammar of Sindhi Language”*. Sindhi Language Authority, Hyderabad Sindhi, Pakistan.
- ANDERSON, STEPHEN. (1985) Inflectional morphology. In Timothy Shopen (ed.), Grammatical categories and the lexicon, volume III of Language typology and syntactic description.
- ANTAL VAN DEN BOSCH and WALTER DAELEMANS. (1999) “Memory-based morphological analysis”, *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, p.285-292, June 20-26, 1999, College Park, Maryland
- ANTWORTH, EVAN L. (1990) “PC-KIMMO: a two-level processor for morphological analysis”. No. 16 in *Occasional publications in academic computing*. Dallas: Summer Institute of Linguistics.
- ANWAR, W., WANG, X. (2006) “A survey of automatic Urdu language processing”. In: *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, Dalian, China, pp. 13–16.
- ATTIA, M. (2006) Accommodating Multiword Expressions in an Arabic LFG Grammar. In Salakoski, Tapio, Filip Ginter, Sampo Pyysalo, Tapio Pahikkala (Eds.): Advances in Natural Language Processing. Vol. 4139, pp. 87–98. Springer-Verlag: Berlin, Heidelberg.
- ATTIA, M. (2008) Handling Arabic morphological and syntactic ambiguities within the LFG framework with a view to machine translation. PhD Dissertation, University of Manchester.
- ATTIA, M., PECINA, P., TORAL, A., TOUNSI, L., & VAN GENABITH, J. (2011) An open-source finite state morphological transducer for modern

- standard Arabic. In Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing (pp. 125-133). Association for Computational Linguistics.
- AUSTIN, P. (2001), Lexical Functional Grammar. In N.J. Smelser and P. Baltes (eds), *International Encyclopedia of the Social and Behavioural Sciences*, 8748-8754. Oxford: Elsevier.
- BACKUS, J. W. (1960) "The syntax and semantics of the proposed International Algebraic Language". In *Proceedings of the International Conference on Information Processing*, UNESCO, Paris, 1959. Oldenbourg, Munich and Butterworth, London, 125--132.
- BARWISE, JON. (1977) "An Introduction to First-Order Logic", in Barwise, Jon, ed. (1982). *Handbook of Mathematical Logic. Studies in Logic and the Foundations of Mathematics*. Amsterdam, NL: North- Holland. ISBN 978-0-444-86388-1.
- BEESLEY, K. R. (1996). "Arabic finite-state morphological analysis and generation". In *COLING-96 Proceedings, volume 1, pages 89 94, Copenhagen. Center for Sprogeteknologi. The 16th International Conference on Computational Linguistics*.
- BEESLEY, K. R. (1998a). Arabic morphological analysis on the Internet. In Proceedings of the 6th International Conference and Exhibition on Multi-lingual Computing.
- BEESLEY, K. R. (1998b). Arabic morphology using only finite-state operations. In Proceedings of the Workshop on Computational Approaches to Semitic languages (pp. 50-57). Association for Computational Linguistics.
- BEESLEY, K. R. and LAURI K. (2003) *Finite State Morphology*. PaloAlto, CA: CSLI Publications.
- BHATT, R., & EMBICK, D. (2003) Causative derivations in Hindi. Manuscript. The University of Massachusetts at Amherst.
- BOGEL, T., BUTT, M., HAUTLI, A. and SULGER, S. (2007) "Developing a finite-state morphological analyzer for Urdu and Hindi". In *Proceedings of the Sixth International Workshop on Finite-State Methods and Natural Language Processing*. Potsdam.
- BOGEL, T. (2010), "Pashto (endo-)clitics in a parallel architecture". In M. Butt and T. H. King, eds., *Proceedings of LFG10*. Stanford: CSLI Publications.

- BRESNAN, JOAN. (2001), *Lexical-Functional Syntax*. Oxford: Blackwell Publishers.
- BRESNAN, JOAN. (ED.) (1982). The mental representation of grammatical relations, The MIT Press, Cambridge, Ma.
- BUCKWALTER, T. (2002) Buckwalter {Arabic} Morphological Analyzer Version 1.0.
- BUTT, M. (1995), *The Structure of Complex Predicates in Urdu*, CSLI Publications, Stanford.
- BUTT, M., & KING, T. H. (2002, AUGUST) Urdu and the Parallel Grammar project. In *Proceedings of the 3rd workshop on Asian language resources and international standardization-Volume 12* (pp. 1-3). Association for Computational Linguistics.
- BUTT, M., & KING, T. H. (2007). Urdu in a parallel grammar development environment. *Language Resources and Evaluation*, 41(2), 191-207.
- BUTT, M., BÖGEL, T., HAUTLI, A. AHMED, T., RAZA, G., and SULGER, S. (2011). "Recent Advances in the Hindi/Urdu LFG Grammar". Talk given at the Workshop on South Asian Syntax and Semantics, University of Massachusetts.
- BUTT, M., SULGER, S., RAHMAN, M. U., & AHMED, T. (2016). Adverb agreement in Urdu, Sindhi and Punjabi. In *Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar* (pp. 140-160).
- BUTT, M., KING, T. H. NIÑO, M. E. and SEGOND, F. (1999) A Grammar Writer's Cookbook. CSLI Publications.
- BUTT, M., KING, T. H., & MAXWELL, J. T. (2003) Complex predicates via restriction. In: *Proceedings of the LFG03 Conference*. CSLI On-line Publications.
- BYBEE, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing.
- CARL POLLARD and IVAN A. SAG. (1994), *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- CARROLL, D. W. (2007) *Psychology of Language*. Cengage Learning.
- CHANOD, J.P. (1994) Finite-state composition of French verb morphology. Technical Report MLTT- 005, Rank Xerox Research Centre- Grenoble Laboratory.

- CHOMSKY, N. (1956) "Three models for the description of language," *PGIT*, 2:3, 113--124.
- CLE. (2016), Sindhi English Dictionary. <http://www.clepk.org/sed1/>. (Accessed 2016).
- CLÉMENT, L. and KINYON, A. (2001), "XLFG: an LFG parsing scheme for French". In *Proceedings of LFG-01, Hong-Kong*.
- COHEN S., YAEL, D. G., and SHULY W. (2003), "Computational Implementation of Non-Concatenative Morphology" in Proceedings of the Workshop on Finite State Methods in Natural Language Processing, Budapest, Hungary, pp. 59-66.
- COLE, J. (2001) Sindhi. In Garry, J. and C. Rubino (eds.), *Facts about the World's Languages: An Encyclopedia of the World's Major Languages, Past and Present*, pp. 647-653.
- COMRIE, B. (1985). *Tense* (Vol. 17). Cambridge University Press.
- CRBLP. (2012), Center for Research on Bangla Language Processing. <http://crblp.bracu.ac.bd/>. (Accessed November 2012)
- CROUCH, D., DALRYMPLE, M., KAPLAN, R., KING, T., MAXWELL, J., & NEWMAN, P. (2011) XLE documentation.
- CSLI ONLINE. (2016), <http://cslipublications.stanford.edu/LFG/> (Accessed August 2016).
- DALRYMPLE, M. (1999), "Lexical-Functional Grammar". In *Rob Wilson and Frank Keil (eds), MIT Encyclopedia of the Cognitive Sciences*. The MIT Press.
- DALRYMPLE, M. (2001) *Lexical-Functional Grammar*. John Wiley & Sons, Ltd.
- DALRYMPLE, M., DYVIK, H., & KING, T. H. (2004) Copular complements: closed or open. In *Proceedings of the LFG04 Conference* (pp. 188-198).
- EVANS, N. (2000). Word classes in the world's languages. *Morphology: A handbook on the inflection and word formation*. Berlin: Mouton de Gruyter.
- FALK, Y. N. (2000). Pivots and the theory of grammatical functions. In *On-line Proceedings of the LFG2000 Conference*. URL csli-publications.stanford.edu/LFG/5/lfg00.html.

- FALK, Y. N. (2001) *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. CSLI. ISBN 1-57586-341-3
- GRIERSON, G. A. (1909). Linguistic Survey of India: Vol. VIII, Part I, Indo-Aryan Family: North Western Group, Specimens of Sindhi and Lahanda. Calcutta: Superintendent of Government Printing, India.
- GÜNTHER, G. and DIETRICH P. (1988) "A finite state approach to German verb morphology", *Proceedings of the 12th conference on Computational linguistics*, p.212-215, August 22-27, Budapest, Hungry.
- HAQ, A., M. (1991) اردو قوائد (Qawaids-i-Urdu), Lahore Academy, Pakistan.
- HAQUE M.N. and KHAN M. (2005) "Parsing Bangla Using LFG: An Introduction", *BRAC University Journal*, 2(1), pp. 105-110.
- HOPCROFT, JOHN E.; ULLMAN, JEFFREY D. (1979) *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley. Chapter 4: Context-Free Grammars, pp. 77–106; Chapter 6: Properties of Context-Free Languages, pp. 125–137
- HUMAYOUN M. and RANTA A. (2010) "Developing Punjabi Morphology, Corpus and Lexicon". *The 24th Pacific Asia conference on Language, Information and Computation (PACLIC24)*.
- HUMAYOUN, M. (2006) Urdu Morphology, Orthography and Lexicon Extraction. Master's Thesis. Department of Computing Science, Chalmers University of Technology.
- HUSSAIN, S. (2004a). Finite State Morphological Analyzer for Urdu. Unpublished MS Thesis, National University of Computer and Emerging Sciences, Lahore, Pakistan.
- HUSSAIN. S and DURRANI N. (2008). *Sindhi in PAN Localization A Study on Collation of Languages from Developing Asia*. CRULP.
- HUSSAIN. S. (2004b), "Urdu Localization Project: Lexicon, MT and TTS", In the Proceedings of Workshop on Computational Approaches to Arabic Script-based Languages, COLING, Geneva, Switzerland.
- IBM NLP. (2012), IBM Natural Language Processing. http://researcher.watson.ibm.com/researcher/view_pic.php?id=147. (Accessed November 2012)
- IIT Kanpur. (2016), Research and Projects. <http://www.cse.iitk.ac.in/users/rmk/proj/proj.html#mt>. (Accessed August 2012).

- IJAZ, M. and HUSSAIN, S. (2007). Corpus Based Urdu Lexicon Development. In *the Proceedings of Conference on Language Technology '07*, University of Peshawar, Peshawar, Pakistan.
- ISA KHADER. (2003), Evaluation of an English LFG-based grammar as error checker. Master's thesis, University of Manchester Institute of Science and Technology.
- JATOI, A. N. (1983). “ਲਿਮੁਲ ਲਿਸਾਵ ਅਤੇ ਸਿਨਡੂ: ਜੁਬਾਵਾ” Linguistics and Sindhi Language, Sindhi Adabi Board. Jamshoro Sindh, Pakistan.
- JOHNSON, C. DOUGLAS. (1972). *Formal Aspects of Phonological Description*. The Hague: Mouton.
- KAMRAN MALIK, M., AHMED, T., SULGER, S., BÖGEL, T., GULZAR, A., RAZA, G., & BUTT, M. (2010). Transliterating Urdu for a Broad-Coverage Urdu/Hindi LFG Grammar. In *LREC 2010, Seventh International Conference on Language Resources and Evaluation* (pp. 2921-2927).
- KAPLAN, R. M., & BUTT, M. (2002). The morphology-syntax interface in LFG. unpublished handout, LFG.
- KAPLAN, RONALD M. and MARTIN KAY. (1994) Regular models of phonological rule systems. *Computational Linguistics* 20 (3):331–378.
- KARTTUNEN, L., & BEESLEY, K. R. (1992) *Two-level rule compiler*. Xerox Corporation. Palo Alto Research Center.
- KENNETH R. BEESLEY and LAURI KARTTUNEN. (2002), *Finite-State Morphology: Xerox Tools and Techniques. Studies in Natural Language Processing*. Cambridge University Press.
- KENNETH W. CHURCH , LISA F. RAU. (1995), Commercial applications of natural language processing, *Communications of the ACM*, v.38 n.11, p.71-79, Nov. 1995.
- KOSKENNIEK K. (1983), Two-level Model for Morphological Analysis, *IJCAI-83, Karlsruhe, BRD*, 683--685.
- KROEGER, P. R. (2005). *Analyzing grammar: An introduction*. Cambridge University Press.
- LAMIA TOUNSI, MOHAMMED ATTIA, and JOSEF VAN GENABITH. (2009), “Parsing arabic using treebank-based lfg resources”. In *Proceedings of the LFG09 Conference*.
- LEECH. G., and WILSON. A. (1996), *EAGLES Recommendations for the Morphosyntactic Annotation of Corpora*. Istituto di Linguistica

Computazionale, Pisa, Italy.

- LEVINE, R. D., and DETMAR M. W. (2006) *Head-Driven Phrase Structure Grammar: Linguistic Approach, Formal Foundations, and Computational Realization*. In Keith Brown. Encyclopedia of Language and Linguistics, Second Edition. Oxford: Elsevier, pp 690-704.
- LFG CONFERENCES. (2016), Lexical Functional Grammar Conferences. <https://www.essex.ac.uk/linguistics/external/LFG/FAQ/conferences.html> (Accessed August, 2016)
- LIONEL C., KIM G. and SYLVAIN K. (2002) *An LFG-type Grammar for German Based on the Topological Model*. CSLI Publications.
- MAHAR J. A. and MEMON G.Q. (2011), Probabilistic Analysis of Sindhi Word Prediction using N-Grams Australian *Journal of Basic and Applied Sciences*, 5(5): 1137-1143
- MAHAR J.A., MEMON G.Q. (2010a). Rule Based Part of Speech Tagging of Sindhi Language, icsap, International Conference on Signal Acquisition and Processing, pp.101-106.
- MAHAR, J.A., MEMON, G.Q. and SHAH, S.H.A. (2010), "WordNet Based Sindhi Text to Speech Synthesis System", *Proceedings of the 2010 Second International Conference on Computer Research and Development*, p.20-24, May 07-10, 2010.
- MAHAR, J.A., MEMON, G.Q. (2010b) Sindhi Part of Speech Tagging System using WordNet, *International Journal of Computer Theory and Engineering* 2(4): 538-545.
- MAHAR, J.A., SHAIKH, H., SOLANGI. A.R. (2011). Comparative Analysis of Rule Based, Syntactic and Semantic Sindhi Parts of Speech Tagging Systems. *International Journal of Academic Research*. Vol. 3. No. 5.
- MANUEL V. F., JORGE G. G. and PILAR A. A. (1996) Finite state morphology and formal verification, *Natural Language Engineering*, v.2 n.4, p.303-304.
- MEGERDOOMIAN, K. (2004), Finite-State Morphological Analysis of Persian. Unpublished manuscript (submitted to Coling 2004).
- MICROSOFT NLP. (2012), Microsoft Natural Language Processing Group. <http://research.microsoft.com/en-us/groups/nlp/>. (Accessed November 2012).
- MOHANAN, T. (1994). *Argument structure in Hindi*. Center for the Study of Language (CSLI).

- MOHAMMED ATTIA, PAVEL PECINA, ANTONIO TORAL, LAMIA TOUNSI, and JOSEF VAN GENABITH. (2011), "An open-source finite state morphological transducer for modern standard Arabic". In *Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing*, pages 125–133, Blois, France. Association for Computational Linguistics.
- MOHRI, M., PEREIRA, F., & RILEY, M. (2002), Weighted finite-state transducers in speech recognition, *Computer Speech and Language*, 16, 69--88.
- MOTLANI Raveesh, Francis M. Tyers, and Dipti M. Sharma. (2016). "A finite-state morphological analyser for sindhi", In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC16)*.
- NLP CENTER. (2012), NLP Center Russia. <http://www.nlpcenter.ru/>. (Accessed Jan 2016).
- NORDLINGER, R. (1998), *Constructive Case: Evidence from Australian Languages*. Stanford: CSLI.
- NORDLINGER, R. AND BRESNAN, J. (2011), *Lexical-Functional Grammar: interactions between morphology and syntax*. In R. Borsley and K. B'orjars, editors, *Non-Transformational Syntax: Formal and Explicit Models of Grammar*. Wiley-Blackwell, Chichester.
- OAD, J. D. (2012), Implementing GF Resource Grammar for Sindhi. Unpublished Master's Thesis. Department of Applied Information Technology Chalmers University of Technology Gothenburg, Sweden.
- PIENEMANN, M. (ED.). (2005) *Cross-linguistic Aspects of Processability Theory* (Vol. 30). John Benjamins Publishing.
- RAHMAN, M. (2009). Sindhi Morphology and Noun Inflections In *the Proceedings of the Conference on Language & Technology CLT09*. CRULP., pp. 74.
- RAHMAN, M. (2010). Towards Sindhi corpus construction. In *Conference on Language and Technology, Lahore, Pakistan*.
- RAHMAN, M., SHAH, A. (2003) Grammar Checking Model for Local Languages. In *proceedings to the SCONEST (Student Conference on Engineering Sciences and Technology)*.
- RAHMAN, M., SHAH, A., and MEMON, R. A. (2007) Partial Word Order Syntax of Urdu/Sindhi and Linear Specification Language. *Journal of*

- Independent Studies and Research (JISR) 5(2): 13-18.
- RANTA, A. (2004), Grammatical Framework: A Type-Theoretical Grammar Formalism, *Journal of Functional Programming* 14 (2): 145–189.
- RIZVI, J. S. M. (2007) *DEVELOPMENT OF ALGORITHMS AND COMPUTATIONAL GRAMMAR FOR URDU* (Doctoral dissertation, Pakistan Institute of Engineering & Applied Sciences, Islamabad).
- ROCHE E and YVES SHABES. (1997), *Finite-State Language Processing*, MIT Press, Cambridge, MA.
- ROSEN, V. (1996), "The LFG Architecture and "Verbless" Syntactic Constructions". In Miriam Butt and Tracy Holloway King. (eds.), *Proceedings of the LFG 96 Conference*. CSLI Publications. <http://csli-publications.stanford.edu/LFG/1/lfg1.html>.
- SAGOT B. and WALTHER G. (2010) "A morphological lexicon for the Persian language". In *Proceedings of the 7th Language Resources and Evaluation Conference (LREC'10)*, Valetta, Malta.
- SCHMIDT, R. L. (1999). *Urdu, an Essential Grammar*. Psychology Press.
- SELLS, PETER. (1985). *Lectures on Contemporary Syntactic Theories*. CSLI Publications.
- SHAIKH, W. B. (2006) “στινδι: o:λι: φo: σαραφ αι~ ναηω” Morphology and Syntax of Sindhi Language. Sindhi Adabi Board.
- SHIEBER, S. M. (1986), An Introduction to Unification- based Approaches to Grammar. CSLI Lecture Notes, CSLI, Stanford, CA.
- SINDHI LANGUAGE AUTHORITY. (2016), Official Website. <http://www.sindhila.org>. (Accessed 2016).
- STANFORD CHINEESE NLP. (2016b), The Stanford Natural Language Processing Group. Chinese Natural Language Processing and Speech Processing. <http://nlp.stanford.edu/projects/chinese-nlp.shtml>. (Accessed August 2016).
- STANFORD NLP. (2016a), The Stanford Natural Language Processing Group. <http://nlp.stanford.edu/>. (Accessed August 2016).
- STEEDMAN, M.J. (1989), A Generative Grammar for Jazz Chord Sequences. *Music Perception* 2 (1): 52–77. JSTOR 40285282.
- STEFANIE D. (2003), Implementing and Documenting Large-scale Grammars -- German LFG. *Ph.D. thesis*, IMS, University of Stuttgart.

- SULGER, S., & VAIDYA, A. (2014). Towards identifying Hindi/Urdu noun templates in support of a large-scale LFG grammar. *COLING 2014*, 1.
- SULGER, S., BUTT, M., KING, T. H., MEURER, P., LACZKÓ, T., RÁKOSI, G., & PATEJUK, A. (2013), ParGramBank: The ParGram Parallel Treebank. In *ACL (1)* (pp. 550-560).
- TRUMPP, E. (1872). *Grammar of the Sindhi Language Compared with the Sanskrit-Prakrit and the Cognate Indian Vernaculars* Ernes Trumpp. Trubner and C.
- URDU PARGRAM. (2016), Urdu Parallel Grammar Project. http://ling.uni-konstanz.de/pages/home/pargram_urdu/ (Accessed, August 2016)
- VIJAY-SHANKER, K. and WEIR, DAVID J. (1994), The Equivalence of Four Extensions of Context-Free Grammars. *Mathematical Systems Theory* 27(6): 511-546.
- VIJAY-SHANKER, K., and JOSHI, A. (1991). *Unification Based Tree Adjoining Grammars*. In Wedekind, J., ed., Unification-based Grammars. Cambridge, Massachusetts: MIT Press.
- WINTNER. S. (2002), "Formal language theory for natural language processing," in *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics.
- XTAG RESEARCH GROUP. (1995), "A Lexicalized Tree Adjoining Grammar for English". Technical report, University of Pennsylvania.
- ZELAL G. and KEMAL O. (1994), "Parsing Turkish text: the lexical functional grammar approach". In *the proceedings of ACL-94*.
- ZAENEN, A. (1989). Nominal Arguments in Dutch and WYSIWYG LFG. Unpublished Manuscript, Xerox PARC.
- ZUHRA, F. T. (2009), A corpus-based finite state morphological analyzer for Pashto. MS (CS) thesis. Peshawar: Department of Computer Science, University of Peshawar.

APPENDIX A

SINDHI PART OF SPEECH TAGSET

S.No.	TAG	Word Class
1	NC	Common Noun
2	NP	Proper Noun
3	NA	Abstract Noun
4	PP1	1st Person Pronoun
5	PP2	2nd Person PN
6	PP3	3rd Person PN
7	PDP	Demonstrative PN Proximate
8	PDR	Demonstrative Pronoun Remote
9	PWh	Wh-Pronoun
10	PRF	Reflexive Pronoun
11	PRL	Relative Pronoun
12	PCR	Co-relative Pronoun
13	PI	Indefinite Pronoun
14	V	Verb
15	Vaux	Passive Intransitive Verb
16	VPPRs	Present Participle
17	VPPst	Past Participle
18	VPFutr	Future Participle
19	VPVNn	Verbal Noun
20	VPConi	Conjunctive Participle
21	AJD	(Characteristic) Adjective
22	AJC	Cardinal
23	AJO	Ordinal
24	AJP	Pronominal Adjectives
25	AJA	Agregative Adjectives
26	AJQ	Quantifier
27	AJF	Fractal
28	AJM	Multiplier
29	AVT	Temporal Adverb
30	AVS	Space Adverb
31	AVM	Manner Adverb
32	AVNeq	Negation Adverb
33	AVQ	Quantity Adverb
34	AVA	Affirmative Adverb
35	AVN	Noun Adverb
36	AVAJ	Adjective Adverb
37	AVP	Pronoun Adverb
38	PP	Post Position
39	PPC	Compound Post Position
40	CC	Coordinate Conjunction
41	CS	Subordinate Conjunction
42	I	Interjection
43	PSxN	Pronominal Suffix with Nouns (Nominal Suffix)
44	PSxV	Verbal Pronominal Suffix
45	PSxP	Postpositional Pronominal Suffix
46	AR	Article
47	R	Residual
48	NU	Numerals
49	PRT	Pre-title
50	POT	Post title
51	SM	Sentence Maker
52	DATE	DATE

APPENDIX B

ROMAN SCRIPT FOR SINDHI

Roman transliteration scheme used includes single lower case letters, single upper case letters and double lower and upper case combinations. Lower case letters are used to represent consonants and short vowels. For example, 'ج', 'گ', 'ب' are represented by 'b', 'g' and 'j' respectively. Short vowels 'ا', 'ي', and 'و' are represented by 'a', 'u' and 'i' respectively. Capital letters are used to represent long vowels and letters with implosive sounds. For example, long vowels 'ا' and 'و' are represented by 'O' and 'U' respectively, and implosives 'ن', 'ب' and 'ج' are represented by 'B', 'D', and 'J' respectively. Other uppercase letters used include rhotic 'ڙ' with 'R' stop 'ٿ' with 'T', and nasal middle 'ن' with 'N' representations. Plosives are represented by two letter combinations. For example, 'پ', 'ڦ', and 'چ' are represented by 'pH', 'bH' and 'jH' respectively. Complete list of letters including vowels along-with their transliterations is given below in Table B-1. Sample transliterated sentences are shown in examples (a) to (e).

(a)	آهيان.	AhiyAN	ويثو	vETHO	ڻ	mEN	گادي	gADI	آئون	AUN				
I am sitting in the car.														
(b)	ھنو.	رہندو	ڻ	ڪراچي	پاءُ	منھنجو	huO	rahaNdO	mEN	karAcHI	bHAu	muhiNjO		
My borther was living in Karachi.														
(c)	ٿيون.	گڏ	اچي	گهر	صدف جي	چوڪريون	سڀئي	tHiyUN	gaDu	acHI	gHaru	jE sadaf	CHOKirYUN	sabHEI
All girls were gathered at sadaf's house.														
(d)	آهن.	رها	نجي	پولڙا	تي	وڻ								

	Ahin	rahiyA	nacHI	bHOliRA	tE	vaNru
Monkeys are dancing on the tree.						
(e)	.هنا کان بدل شينهن	جانور jAnvar	جا jA	جهنگ jHaNg		
Animals of the jungle were afraid of the loin.						

Table B-1: Partially Used LFG Rules.

letter	Roman	Letter	Roman	Letter	Roman	Letter	Roman
ا	a	۽	CH	ٻ	t	ڻ	n, N
ٻ	b	ڦ	H	ڙ	z	ڦ	Nr
ڦ	B	ڦ	KH	ڻ	a	ڻ	v
ڦ	bH	ڏ	D	ڻ	G	ڻ	a
ڌ	t	ڌ	dH	ڻ	f	ڻ	h
ٿ	tH	ڻ	D	ڙ	pH	ڻ	y
ٿ	T	ڦ	dd	ڦ	q		
ٿ	TH	ڦ	DH	ڪ	k	ڻ	A
ٿ	s	ر	R	ڪ	kH	اي	E
ٻ	p	ڙ	R	ڱ	g	اي	I
ڇ	j	ڙ	z	ڳ	G	او	O
ڇ	J	س	S	ڳ	gH	او	U
ڇ	jH	ش	sH	ڱ	Ng	اُ	u
ڇ	NG	ص	S	ڻ	I	ا	i, e
ڇ	cH	ض	z	ڻ	m	ڻ	a