

# Finite State Morphology and Sindhi Noun Inflections

Mutee U Rahman, Mohammad Iqbal Bhatti

Department of Computer Science, Isra University,  
Hala Road, Hyderabad Sindh 71000, Pakistan  
muteerahman@gmail.com, iqbalbhatti@isra.edu.pk

**Abstract.** Sindhi is a morphologically rich language. Morphological construction include inflections and derivations. Sindhi morphology becomes more complex due to primary and secondary word types which are further divided into simple, complex and compound words. Sindhi nouns are marked by number gender and case. Finite state transducers (FSTs) quite reasonably represent the inflectional morphology of Sindhi nouns. The paper investigates Sindhi noun inflection rules and defines equivalent computational rules to be used by FSTs; corresponding FSTs are also given.

**Keywords.** Sindhi, morphology, noun inflections, two-level morphology, finite state morphology.

## 1 Introduction

Morphology deals with word formation rules in a language. Word structures of a language are defined by its morphological constructions. Morphology defines that how smaller meaning bearing units called morphemes are combined to make larger meaning bearing units of a language called words. Morphology also deals with word formation by variations in already existing words. The morphological changes are mostly done by suffix addition, subtraction and replacement phenomenon. In few words morphology can be defined as syntax of word formation.

Models for computational analysis of morphology always remained challenge for computational linguists until early 1980's when 4Ks\* discovered the two level morphology (Kaplan, R. M. and M. Kay. 1981) the first general model for morphologically complex languages. This two level morphology represents a word at lexical level and surface level. Morphotactics or morpheme ordering model is used in between these two levels to incorporate morphological changes. These morphotactics are implemented as Finite State Transducers (FSTs).

Sindhi is one of the major languages of Pakistan and is spoken by approximately 40 million people (Sindhi Language Authority. 2009). Sindhi is an example of morphologically complex language. Nouns are marked by number, gender and case. Two level morphology can be used to model Sindhi noun inflections.

Subsequent sections discuss Finite State Morphology, Sindhi morphological constructions, Sindhi noun inflections and role of finite state transducers in their computation. Section 2 discusses Sindhi noun inflections in detail. In section 3 finite state transducers for different noun inflection types in Sindhi are presented. Conclusions are discussed in section 4 and finally references are given in section 5. IPA Sindhi transliteration along-with Arabic script is also given.

### 1.1 Finite State Morphology

Finite state transducers play an important role in language processing applications (Beesley,

---

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

Kenneth R. and Lauri Karttunen, 2003) and computational studies of morphologically complex languages. Morphotactics (morpheme ordering rules) and orthographical rules (spelling rules) are represented by finite state transducers. Efficient morphological parsers can be implemented by combining these finite automata and computational lexicon (repository of words).

Finite state transducers convert/translate lexical level constructs to surface level words by applying morphotactics and orthographical rules. Their reversible nature makes reverse conversion/translation possible. This two level (lexical and surface) morphology plays crucial role in implementation of morphological analyzers for natural languages.

## 1.2 Morphological Construction in Sindhi

Sindhi is an example of Indo-Aryan language with rich inflectional and derivational morphology (Mutee-U-Rahman, 2009). Sindhi morphological constructions include derivational and inflectional morphology with addition, subtraction and replacement methods. Sindhi words are divided into primary and secondary word types. Secondary words are further divided into compound and complex words. Following sections discuss Sindhi word types and their morphological construction in detail.

**Sindhi Words.** Sindhi words are divided into two types primary or simple words and secondary words (Jatoi Ali Nawaz, 1983). Primary words (also known as minimum free forms) are not further divisible (Khubchandani, Lachman M. 2003). For example *fa:na* چاڻ (knowledge) and *rasto* رستو (path or way) are examples of primary words.

Secondary words are divided into complex and compound words. Complex words are formed by combining affixes with primary words. For example primary word *fa:na* چاڻ when combined with prefix *an* اڻ (negation) becomes a complex word *anfa:na* اڻچاڻ (layman). Same word when combined with suffix *o*: او becomes *fa:no*: چاڻو (scholar). Compound words are combinations of two or more simple words. Their prefixes and suffixes are actually free form morphemes. *dʒ<sup>h</sup>aqbilu*: جهنگڀلو (wild cat) which is formed by combining two free form morphemes *dʒ<sup>h</sup>aga* جهنگ (forest) and *bilu*: ڀلو (cat); and *hat<sup>h</sup>kaṛi*: هٽڪڙي (manacle) which is formed by *hat<sup>h</sup>a* هٽ (hand) and *kaṛi*: ڪڙي (ring) are examples of compound words. Words in Sindhi always end in a vowel (Sheikh Wahid Bakhsh, 2006). These endings not only help in identifying the gender in case of nouns but change in them can cause a different word class or derivational morphology. Words can have following vowel endings.

Sindhi:	اَ	آ	إِ	إِي	أُ	أُو	او	أَو	اي	آي
IPA:	a	a:	e	e:	u	u:	o	o:	i	i:

**Morphological Construction in Sindhi.** Sindhi is a polymorphemic language. Sindhi morphological constructions include derivational and inflectional forms. Sindhi derivations take place when word stem is combined with a grammatical morpheme usually resulting in a different class word. For example the adjective *buk<sup>h</sup>yo*: بڪيو (hungry) is derived from noun *buk<sup>h</sup>a* بڪ (hunger) when suffix *yo*: يو 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 *pok<sup>h</sup>a* پوڪ (crop) is derived from verb *pok<sup>h</sup>e* پوڪ (sow) by changing of last vowel “e” to “a”.

Sindhi inflectional morphology takes place by combining a word stem with a morpheme resulting in word of same class which performs same syntactic function as the original stem. Inflections are caused by change in gender, number case or tense. Sindhi nouns are marked by number, gender and case.

## 2 Sindhi Noun Inflections

Sindhi nouns are divided in two major categories Concrete Nouns and Abstract Nouns. Concrete nouns are further divided into Common Nouns and Proper Nouns (Baig, M. Q, 2006). As discussed in section 1.2.1 Sindhi words always end in a vowel so is the case with nouns; these endings also identify the gender of a noun. Following sections discuss noun inflections with respect to gender, number and case.

## 2.1 Gender

Sindhi nouns have two genders masculine and feminine. This gender classification is for animate and non-animate nouns. For example g<sup>h</sup>ara گهر (house) in Sindhi is masculine and hotale هوٽل (hotel) is feminine. Gender of non-animate nouns is mostly defined artificially and usually smaller things are considered feminine and larger ones are masculine (there are some exceptions shown in Table 1). As discussed above gender of nouns is mostly identified by last vowel/diacritic sound. Feminine nouns mostly end in اَ, اِ, اِي (a, a:, e, e:) endings and masculine nouns usually end with اُ, اُو, او (u, u: ū:, o:) endings; there are some exceptions like the common noun pak<sup>h</sup>e: پڪي (bird) is masculine with e: ending. Table 1 shows examples of masculine and feminine nouns.

## 2.2 Number

Like English, Sindhi nouns also have two numbers Singular and Plural. Number inflections depend on the gender of noun and ending vowels/diacritics. Number inflections in feminine and masculine nouns take place differently. Table 2 shows some examples of feminine and masculine nouns along-with their number inflections.

**Table 1.** Sindhi Masculine and Feminine Nouns.

	Word	Ending	Gender	English Meaning
Animate Nouns	zāla زال	اَ (a)	F	Wife
	dunyā: دنيا	اَ (a:)	F	World
	rāte رات	اِ (e)	F	Night
	tj <sup>h</sup> okri: چوڪري	اِي (e:)	F	Girl
	bāru پار	اُ (u)	M	Child
	vitj <sup>h</sup> ū: وچون	اُو (u:)	M	Scorpion
	tj <sup>h</sup> okro: چوڪرو	او (o:)	M	Boy
Non-animate Nouns	dara در	اَ (a)	M	Door
	dari: دري	اِي (i:)	F	Window
	tānbu: تنبو	اُو (u:)	M	Tent
	d <sup>h</sup> arti: ڌرتي (Exception, bigger non-animate but feminine)	اِي (i:)	F	Earth

## 2.3 Case

Linguists define five different cases in Sindhi case system which are given below:

- (i) Nominative
- (ii) Accusative-Dative
- (iii) Postpositional
- (iv) Genitive
- (v) Vocative

Nouns are not inflected in nominative case and remain in their original form. In accusative-

dative, postpositional and genitive cases nouns are inflected and their inflected forms remain same in these cases. This same inflected form in these three cases is known as Oblique Case. Examples of nominative and oblique forms of noun  $tʃ^hokiro$ : چوڪرو (boy) are shown in Table 3.

**Table 2.** Number inflections in feminine and masculine nouns.

Gender	Singular Noun		Plural Noun		Ending Vowel
	Sindhi	English	Sindhi	English	
Feminine	zāla زال	Wife	zālū: زالون	Wives	اَ (a)
	$tʃ^hokri$ : چوڪري	Girl	$tʃ^hokirū$ : چوڪريون	Girls	اي (e:)
	hava: هوا	Wind	hava:ū: هوائون	Winds	اَ (a:)
Masculine	$tʃ^hokiro$ : چوڪرو	Boy	$tʃ^hokira$ : چوڪرا	Boys	او (o:)
	putu پٽ	Son	puta پٽ	Sons	اُ (u)
	$pak^hi$ : پڪي	Bird	$pak^hi$ : پڪي	Birds	اي (i:)

**Table 3.** Nominative and oblique forms of noun  $tʃ^hokiro$ .

Gender	Number	Nominative	Oblique
Feminine	Singular	$tʃ^hokiro$ : چوڪرو	$tʃ^hokire$ : چوڪري
	Plural	$tʃ^hokira$ : چوڪرا	$tʃ^hokirana$ : چوڪرن
Masculine	Singular	$tʃ^hokiri$ : چوڪري	$tʃ^hokiri:a$ : چوڪريءَ
	Plural	$tʃ^hokirū$ : چوڪريون	$tʃ^hokiruna$ : چوڪرين

In Sindhi case system vocative case is formed by prefixing an interjection before nominative. For example  $ae: dosta$  (o friend) and  $aʃe: dosta$  (oh friend). Table 4 shows some examples of vocative case.

**Table 4.** Sindhi vocative case examples.

Number	Gender	Nominative	Meaning	Vocative	Meaning
Singular	M	$\delta a:ru$ ٻارُ	Child	$\delta a:ra$ ٻارَ	O Child!
		$va:d^ho$ واڻو	Carpenter	$va:d^ha$ واڻا	O Carpenter!
	F	$b^hiŋu$ ڀيڻُ	Sister	$b^hiŋa$ ڀيڻَ	O Sister!
Plural	M	$\delta a:ra$ ٻارَ	Children	$\delta a:rao$ ٻارو	O Children!
		$va:d^ha$ واڻا	Carpenters	$va:d^hao$ واڻو	O Carpenters!
	F	$b^hiŋū$ : ڀيڻون	Sisters	$b^hiŋao$ : ڀيڻو	O Sisters!

### 3 Finite State Transducers and Sindhi Noun Inflections

Finite state transducers (FSTs) are capable enough to model Sindhi noun inflections. Two level morphology along-with morphotactics and orthography rules can be used to represent inflections in Sindhi nouns. Following sections discuss gender, number and case inflection rules for Sindhi nouns and corresponding finite state models.

### 3.1 Finite State Transducers for Gender Inflections in Sindhi Nouns

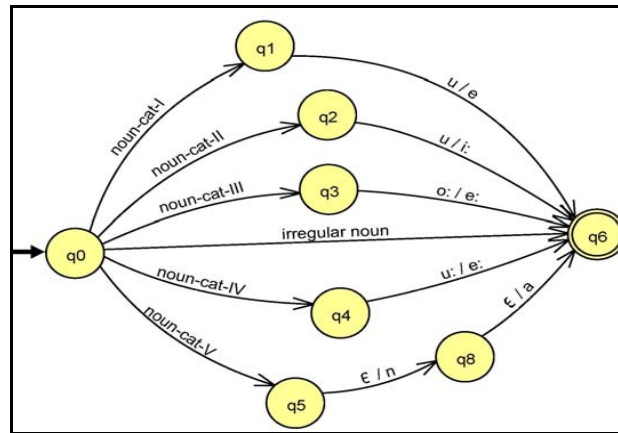
As discussed in previous sections gender in Sindhi is mostly identified by ending vowel sound; inflection therefore takes place by last vowel change. This change may occur differently with same endings. Therefore nouns are categorized to define inflection rules. Table 5 shows gender inflection rules in Chomsky and Halle (Chomsky, Noam , Morris Halle. 1968) notation for Sindhi nouns of different categories. For example the rule “ $u \rightarrow e / \epsilon \_ \#$ ” says that translate “u” to “e” whenever “u” occurs between  $\epsilon$  (null symbol) and # (word boundary); which ultimately means that if “u” occurs in the end of a noun it will be replaced by “e”. Figure 1 shows corresponding finite state transducer.

**Table 5.** Gender inflection rules for Sindhi nouns.

Noun		Rule	Category
Masculine	Feminine		
gadahu گڏھ	gadahe گڏھ	$u \rightarrow e / \epsilon \_ \#$	I
ḡakaru بڪر	ḡakari بڪري	$u \rightarrow i: / \epsilon \_ \#$	II
ḡelo: ٻلو	ḡele: ٻلي	$o: \rightarrow i: / \epsilon \_ \#$	III
nanu نانو	nane ناني	$u: \rightarrow e: / \epsilon \_ \#$	IV
uṭ <sup>h</sup> e اٺ	uṭ <sup>h</sup> ṇa اٺڻ	$\epsilon \rightarrow \eta a / \epsilon \_ \#$	V

FST of Figure 1 recognizes or accepts the nouns of different categories until their last vowel occurrence which is then translated to appropriate vowel to reflect the gender inflection. The conventional colon (:) mark for mapping input and output in FST is replaced by slash (/) to avoid confusions as colon (:) is part of different IPA symbols.

There are some examples of irregular gender inflections for which rules are not defined. For example feminine of muṛsa مڙس (Husband) is za:la زال (Wife) and feminine of puṭa پٽ (son) is d<sup>h</sup>i:a ڌيءَ (daughter).



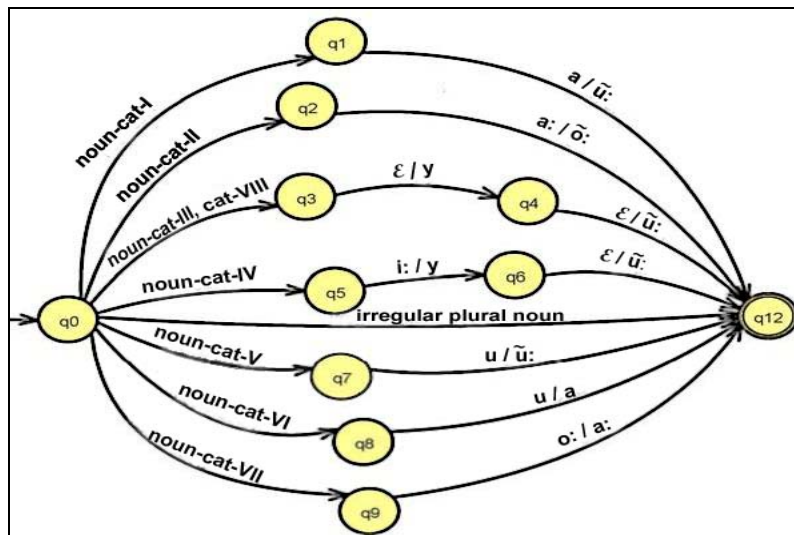
**Figure 1.** FST for gender inflections in Sindhi nouns.

### 3.2 Finite State Transducers for Number Inflections in Sindhi Nouns

Number inflections in Sindhi nouns take place according to gender and last vowel ending. Table 6 shows rules for different noun categories according to their gender and vowel endings. Finite state transducer for rules of Table 6 is shown in Figure 2.

**Table 6.** Number inflection rules for Sindhi nouns.

Gender	Singular	English	Plural	Rule	Category
Feminine	me:za مِزَ	Table	me:zū: مِزُون	$a \rightarrow \bar{u}: / \varepsilon\_\_\#$	I
	hava: هوا	Wind/air	havaō: هَوَانُون	$a: \rightarrow \bar{o}: / \varepsilon\_\_\#$	II
	ra:te رات	Night	ra:teyū: رَاتِيُون	$\varepsilon \rightarrow y\bar{u}: / \varepsilon\_\_\#$	III
	tʃ <sup>h</sup> okiri: چوڪري	Girl	tʃ <sup>h</sup> okiryū: چوڪريُون	$i: \rightarrow y\bar{u}: / \varepsilon\_\_\#$	IV
	b <sup>h</sup> iṅu پيڻ	Sister	bhiṅū: پيڻُون	$u \rightarrow \bar{u}: / \varepsilon\_\_\#$	V
Masculine	kita:bu ڪتابُ	Book	kita:ba ڪتابَ	$u \rightarrow a / \varepsilon\_\_\#$	VI
	hafto: هفتو	Week	hafta: هفتا	$o: \rightarrow a: / \varepsilon\_\_\#$	VII
	se:t <sup>h</sup> e سيٺ	Businessman	se:t <sup>h</sup> eyū: سيٺِيُون	Category III Rule	VIII
	pakhi: پڪي	Bird	pakhi: پڪي	$i: \rightarrow i: / \varepsilon\_\_\#$ i rregular	IX

**Figure 2.** FST for number inflections in Sindhi.

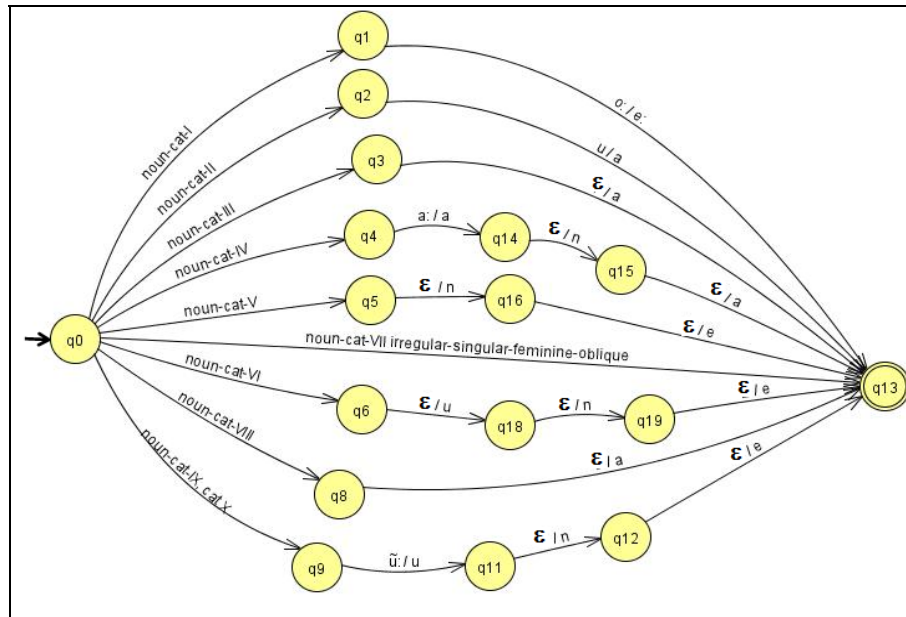
### 3.3 Finite State Transducers for Oblique and Vocative Case Inflections

As discussed in section 2.3 nouns are not inflected in nominative case but in accusative-dative, postpositional and genitive cases their inflected form remain same and is known as oblique case. Inflection rules for oblique case are given in Table 7. In noun category VII irregular inflection takes place and noun remains in its original form. The corresponding finite state transducer is given in Figure 3.

When interjection is prefixed before nominative the vocative case is formed. Sindhi nouns are inflected in vocative case. Table 8 shows vocative case noun inflection rules for different noun categories along-with their nominative and vocative forms. Noun category IV is example of irregular vocative inflection in which noun remains in its original form. Figure 4 shows FST for vocative case inflections.

**Table 7.** Sindhi oblique case noun inflection rules.

Gender	Number	Nominative	Oblique	Rule	Category
Masculine	Singular	tʃ <sup>h</sup> okiro:	tʃ <sup>h</sup> okire:	o: → e: / ε <sub>...</sub> #	I
		putu	puta	u → a / ε <sub>...</sub> #	II
		hat <sup>h</sup> i:	hat <sup>h</sup> i:a	ε → a / ε <sub>...</sub> #	III
	Plural	tʃ <sup>h</sup> okira:	tʃ <sup>h</sup> okirana	a: → ana / ε <sub>...</sub> #	IV
		puta	putane	ε → ne / ε <sub>...</sub> #	V
		hat <sup>h</sup> i:	hat <sup>h</sup> i:une	ε → une / ε <sub>...</sub> #	VI
Feminine	Singular	k <sup>h</sup> aʃa	k <sup>h</sup> aʃa	a,e,u,a: → a,e,u,a: / ε <sub>...</sub> # irregular singular feminine oblique	VII
		b <sup>h</sup> ite	b <sup>h</sup> ite		
		ma:u	ma:u		
		hava:	hava:		
		tʃ <sup>h</sup> okiri:	tʃ <sup>h</sup> okir:a	ε → a / ε <sub>...</sub> #	
	Plural	k <sup>h</sup> aʃū:	k <sup>h</sup> aʃune	ū: → une / ε <sub>...</sub> #	IX
		b <sup>h</sup> ityū:	b <sup>h</sup> ityune:	ū: → une / ε <sub>...</sub> #	X



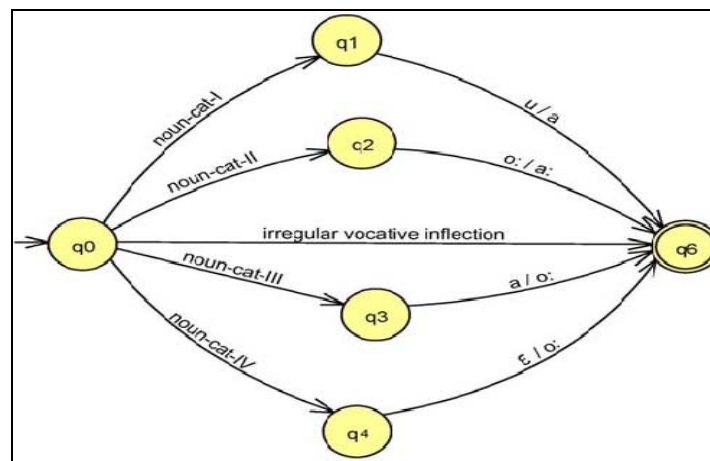
**Figure 3.** FST for oblique case noun inflections in Sindhi.

#### 4 Conclusion

Finite state transducers play important role in computational treatment of Sindhi nouns. They represent morphotactics of Sindhi noun inflections quite reasonably and provide computational basis for automatic morphological processing of Sindhi nouns. Gender, number and case inflection rules and their corresponding FSTs given in section 3 can be further extended to complete morphological analyzers by incorporating Sindhi computational lexicon. Lexicon to surface level generations can be achieved by applying FSTs presented. Surface to lexicon level morphological parsing can also be achieved automatically once lexicon to surface level mapping is worked out and properly modeled by using inflection rules and FSTs. Reversible nature of FSTs makes it possible to execute the whole process of mapping lexicon to surface levels in reverse order. Therefore noun inflection FSTs discussed here are not only useful in inflected word formation but also in morphological analysis/parsing of Sindhi nouns.

**Table 8.** Sindhi noun inflections in vocative case.

Gender	Number	Nominative	Vocative	Rule	Category
Masculine	Singular	ḡa:ru	ḡa:ra	$u \rightarrow a / \varepsilon_{\sim}\#$	I
		ṽad <sup>h</sup> o:	ṽad <sup>h</sup> a:	$o: \rightarrow a: / \varepsilon_{\sim}\#$	II
	Plural	ḡa:ra	ḡa:ro:	$a \rightarrow o: / \varepsilon_{\sim}\#$	III
Feminine	Singular	b <sup>h</sup> iṅu,	b <sup>h</sup> iṅu	$a, a:, i:, u \rightarrow$ $a, a:, i:, u / \varepsilon_{\sim}\#$ irregular	IV
		ma:si:	ma:si:		
		dua:	dua:		
	Plural	b <sup>h</sup> iṅu	b <sup>h</sup> iṅuo:	$\varepsilon \rightarrow o: / \varepsilon_{\sim}\#$	V
		ma:si:	ma:si:o:		
		dua:	dua:o:		

**Figure 4.** FST for vocative case noun inflections in Sindhi.

## References

- Baig, M. Q. 2006. "sind<sup>h</sup>i: viyā karaṇa" Sindhi Grammar Part II, Sindhi Adabi Board. Jamshoro Sindh, Pakistan.
- Beesley, Kenneth R. and Lauri Karttunen. 2003. Finite State Morphology. PaloAlto, CA: CSLI Publications.
- Chomsky, Noam, Morris Halle. 1968. The Sound Pattern of English. New York: Harper and Row.
- Jatoi Ali Nawaz. 1983. "ilimul lisān aī sind<sup>h</sup>i: zubāna" Linguistics and Sindhi Language, Sindhi Adabi Board. Jamshoro Sindh, Pakistan.
- Kaplan, R. M. and M. Kay. 1981. Phonological rules and finite-state transducers {Abstract}. Linguistic Society of America Meeting Handbook. Fifty-sixth Annual Meeting, December 27 - 30, New York.
- Khubchandani, Lachman M. 2003. "Sindhi." George Cardona & Dhanesh Jain (eds.), The Indo-Aryan Languages. London / New York: Routledge (Routledge Language Family Series). 622-58
- Mutee-U-Rahman. 2009. Sindhi Morphology and Noun Inflections. In proceedings of Conference on Language and Technology CLT09. Pp 74-81. FAST National University. Lahore Pakistan.
- Sheikh Wahid Bakhsh. 2006. "sind<sup>h</sup>i: ḡo:li: dzo: saraf aī nahava" The Morphology and Syntax of Sindhi Language, Sindhi Adabi Board. Jamshoro Sindh, Pakistan.
- Sindhi Language Authority, Website <http://sindhila.org>. Accessed on March 2010.