

Phonotactic Distribution of Consonantal Sounds in Urdu

Abstract: *Phonotactic distribution of any language represents all possible phoneme combinations allowed by that language. In this paper, we try to describe the phonotactic distribution of Urdu language with respect to consonantal phonemes. We have also presented possible Urdu syllable templates with complex consonantal cluster.*

Keywords: *Phonotactics, coda consonants, sonority, syllable, templates, Phonotactic distribution*

1. INTRODUCTION

Phonotactic distribution is a basic set of rules of a language on which further phonotactic and phonological descriptions are made. This paper was written at the conclusion of our study to find all possible phonotactic distribution of consonantal segments of Urdu. The paper is also indented to give all possible syllable templates, which have complex consonantal clusters within them. However, the study is limited to coda consonants.

2. LITERATURE REVIEW

Urdu is classified as Indo-Aryan language, which is a major sub branch of Indo-European. Urdu is spoken by at least 60 million people in more than 10 countries (Hussain, 1997).

The distinctive sound-units or phones of a language are the basic building blocks, which combine to form clusters known as syllables. There are certain rules to form these syllables e.g., Maximal Onset Principal (MOP) (Goldsmith, 1990, pg. 137) or Sonority Sequence Principal (SSP) etc.) (Goldsmith, 1990, pg. 105 and 107).

Every syllable has two major parts: onset and rhyme. Further rhyme has two components nucleus and coda. Nucleus always constitutes one or two vowels. Both onset and coda can have zero or more consonants (Goldsmith, 1990, pg. 108-109).

2.1. Sonority Hierarchy

According to SSP a syllable must have rising sonority till the nucleus and falling from there onward (Goldsmith, 1990). The sonority hierarchy of groups of alphabets on the basis of manner (from high to low) is as follows:

- Vowels
 - Low
 - Mid
 - High
- Glides
- Liquids
- Nasals
- Obstruents

- Fricatives
- Affricates
- Stops

Sometimes, languages do add extra syllabic material at the end of word that may violate SSP (Goldsmith, 1990, pg. 107-108).

2.2. Syllable Templates

As suggested by (Hussain, 1997), Urdu can have complex onset and coda clusters and proposed following possible syllable template with consonant clusters:

CCV, CCVC, CCVCC, CVCC, VCC, CCVV, CCVVC, CCVVCC, CVVCC, VVCC. Where “V” represents a short vowel and “C” represents a consonant. The “VV” cluster represents long vowel, as two short vowels cannot come consecutively (Bokhari, 1985).

2.3. Phonotactics

Phonotactics or Phonotactic Distribution is the study of determining the sequences of segments allowed in a language and those, which are not allowed in that language, are called phonotactic constraints.

According to (Hussain, 1997), Urdu allows voiceless fricatives and nasals the first position in CC cluster of coda of a syllable and the second consonant is a stop.

As claimed by (Gazali, 2002), first consonant in CC coda cluster can also be one of the liquids, nasals, voiced fricatives [] and [z] or the stops [b], [k] and [t̪]. These configurations are limited for CVCC template and not for CVVCC template. Only the voiceless fricatives exist in both syllable templates. Furthermore, the last consonant in CVVCC cluster is extra-syllabic. The syllable template VCC also exists in Urdu.

(Gazali, 2002), (Akram, 2002) and (Nazar, 2002) claim that the syllable template VVCC is totally prohibited in Urdu. But (Nayyar, 2002) has claimed that Urdu also has the syllable template VVCC.

3. METHODOLOGY

To find the complete phonotactic distribution of Urdu, we needed to analyze almost all its words. So we completely covered the “Standard Twentieth Century Dictionary” (Merg), which covers more than 50,000 words of Urdu. To verify the syllabification of these words (Ferozsons, 1967) was consulted.

According to (Saleem, et.al. 2002), Urdu has 43 consonantal sounds. We left two consonantal sounds [] and [ɟ] that are not commonly used in Urdu dialect spoken in Lahore and its surrounding areas. Moreover, (Khalid, 2002) claims that the fricative [v] and the glide

[w] are indistinguishable by the Urdu native speakers and they tend towards a pronunciation in the middle of the approximant [w], the fricative [v] and the vowel [u]. So, in this paper, 41 consonantal sounds are used, which are listed as under in Table 1.

Table 1 Urdu Consonants

Groups	Elements
Stops	[b], [b̥], [p], [p̥], [t̪], [d̪], [t], [d], [ʈ], [ʈ̥], [ʈ̣], [ʈ̣̥], [ʈ̣̣], [k], [k̥], [ʃ], [ʃ̥]
Fricatives	[f], [s], [z], [ʃ], [ʒ], [x], [h], [h̥]
Affricates	[tʃ], [dʒ], [tʃ̥], [dʒ̥]
Nasals	[m], [m̥], [n], [n̥], [ŋ]
Liquids	[l], [l̥]
Trill	[r], [r̥]
Flap	[ɾ], [ɾ̥]
Glides	[w], [ɰ]

As mentioned by (Akram, 2002), Urdu has at most three consecutive consonants in a word, which is broken by syllable boundary according to MOP followed by SSP. So, an Urdu word has at most two consonantal segments in coda or onset. Therefore, we build a 2x2 matrix that has all the consonantal phones of Urdu in a column and a row. Every syllable having complex consonant cluster fills an entry of this matrix, e.g. the word [ɾab̪t̪] has a complex coda [ɾt̪], so it will occupy place corresponding to [b] in the column and [t̪] in the row. The collected data during the research is presented in the Appendix.

3.1. Phonotactic Distribution

Every word from the dictionary was at first syllabified first. For proper syllabification, help was taken from native speakers including the writers of this paper. Any syllable having consonantal cluster were noted in the matrix. Separate cluster information is noted for onsets and codas.

3.2. Syllable Templates

After syllabification of dictionary words, all known syllable templates were verified from the matrix formed and the templates, which were not found in any word, were left apart.

3.3. Sonority Sequence

Sonority sequence was determined by finding any syllable having consonant cluster of same group (table 1) not present on word boundary, because on word boundary we may have extra syllabic material.

4. RESULTS

The data was collected for both coda and onset positioned consonant clusters. The following results of the research are only for complex consonant cluster at coda position. Complete data in a simple and unambiguous form so that future work can be done over the same data without exerting efforts on data collection.

Table 2 enlists consonantal sounds in the left column of the table that allow the consonantal sounds in the right column to come after them.

Table 2 After Consonantal Sounds

[ʈ]	[b, t, d, , k, f, w, s, x, h, m]
[ɾ]	[b, t, d, , k, , f, s, z, , x, , h, t, d , m , n]
[b]	[t, d, s, z, , l, r, n]
[f]	[t, d, , k, s, z, , r, n]
[m]	[b , p, t, s, z, l, r, n]
[p]	[t]
[s]	[b, p, t, d, k, f, l, r, m , n]
[z]	[b, d, k, f, l, r, m , n]
[ʃ]	[b, t, d, k, f, r, m , n]
[ʒ]	[d, s, z, l, r, m]
[x]	[t, s, z, , l, r]
[h]	[w , z, m]
[h̥]	[b, t, d, f, w , s, z, , l, r, m , n,]
[ʃ̥]	[b, k, f, l, r, m , n]
[ʒ̥]	[s, l, r, h]
[x̥]	[b, t, d, f, z, s, , l, r, m , n]
[ḥ]	[ʃ̣, ẓ, l, r, m]
[ḥ̥]	[t, t, d, d , t, t, d, k, k , f, s, z, h, t, d , t , d]
[x̣]	[t, s, z, , l, r]
[ʃ̣̥]	[w , z, m]
[ḥ̥]	[b, t, d, f, w , s, z, , l, r, m , n,]
[ʃ̣̣]	[b, k, f, l, r, m , n]
[ʒ̣̥]	[s, l, r, h]
[x̣̥]	[b, t, d, f, z, s, , l, r, m , n]
[ḥ̣]	[ʃ̣̣, ẓ̣, l, r, m]

Table 3 enlists the consonantal sounds in the left column of the table that allow the consonantal sounds in the right column to come before them.

Table 3 Before Consonantal Sounds

[ʈ]	[b, t, d, k, f, s, z, x, h, d , m]
[ɾ]	[b, t, d, k, f, s, z, , x, h, d , m]
[b]	[t, k, s, z, , h, l, r]
[f]	[t, k, s, z, , h, l, r, n]
[m]	[b , p, t, k, s, z, , , h, d , l, r]
[p]	[s, m]
[b̥]	[m]
[s]	[b, d, k, f, x, , h, d , l, r, m , n]
[z]	[b, k, f, x, h, d , r, m , n]
[ʃ]	[b, k, f, x, h, r]
[ʒ]	[ʃ̣, ẓ, l, r, m]
[x]	[r, n]
[h]	[r, n]
[ʃ̥]	[n]
[ʒ̥]	[n]
[x̥]	[b, t, k, f, s, z, , h, r, m]
[h̥]	[h]
[ʃ̣̥]	[s, l, r]
[ʒ̣̥]	[x]
[x̣̥]	[d, s, l, n]
[ḥ̥]	[n, l, r]
[ʃ̣̣̥]	[n]
[ʒ̣̣̥]	[n]
[x̣̣̥]	[b, p, k, f, s, , x, h, l, r, m , n]
[ḥ̣̥]	[n]
[ʃ̣̣̣̥]	[n]
[ʒ̣̣̣̥]	[n]
[x̣̣̣̥]	[b, k, f, s, z, , h, d , l, r, n]
[ḥ̣̣̥]	[n]
[ʃ̣̣̣̣̥]	[b, t, f, s, z, , l, r, n]

[k]	[n]
[]	[r]
[]	
[w]	[r, h, l]
[j]	

5. DISCUSSION

5.1. Syllable Templates with Consonant Clusters

Urdu allows at most two consonantal segments in coda of a syllable. Thus, the possible syllable templates with complex consonantal cluster in coda position are shown in Table 4.

Table 4 Urdu Syllable Templates

Syllable Template	Examples
VCC	[k], [rz]
CVCC	[w kt], [p nd r]
CVVCC	[s p], [nd]

The data shows that a syllable with missing onset and a consonant cluster in coda (i.e. VVCC) does not allow long vowel at its nucleus. The results match the results of (Gazali, 2002), (Akram, 2002) and (Nazar, 2002) that have claimed that Urdu does not have syllable template VVCC but disagrees with the results of (Nayyar, 2002).

Moreover, Urdu does allow CC cluster at the onset position of syllable. When some words are spoken in hurry, some syllable segments are skipped from those words and resulting in CC cluster at onset position e.g. [d r xt] becomes [dr xt]. Beyond this, there are some words that originally have CC cluster in onset position e.g. [l k.m j] and [kya]. This paper does not deal with such examples of complex onset. This topic needs to be studied separately.

5.2. Phonotactic Distribution of Urdu Consonants

On the basis of results, a proposed classification of segments for coda consonant cluster in Urdu is given in Table 5. These categorical groups of segments (table 1) are self-exclusive i.e. any two members of a group do not come one after the other. Only the non-labial stops are the exceptions that should be further classified. For example, an affricate is never followed by another affricate and a labial never follows any other labial.

Table 5 Categorical Groups for Coda Consonant Cluster

Groups	Elements
Labials Stops	[b], [b], [p], [p], [m], [m], [f]
Glides	[w], [j]
Liquids/Trill/Flap	[l], [l], [z], [r], [], []
Coronal Fricatives	[s], [z], [], []
Non-Coronal Fricatives	[x], [], [h]
Affricates	[t], [t], [d], [d]
Non-Labial Nasals	[n], [n], []
Non-Labial Stops	[t], [d], [t], [d], [], [], [], [], [k], [k], [], []

The results also showed that some consonantal segments are very rare in Urdu. The list includes [l], [r], [], [], [] and [n]. The segments that never occur in any consonant cluster are [], [], [j], [l], [r], [], [], [m] and [n]. Other less frequent segments are [b], [p], [p], [], [], [], [], [], [], [], [k] and [].

When stops come at first position in CC cluster of the coda, it can be followed by a stop (nasal or non-nasal), a coronal fricative, a liquid or a trill, as illustrated in Figure 1, e.g., in the words [w kt], [s bz] and [d br]. These CC clusters always occur at word end and the second consonant can be referred to as extra syllabic material, as otherwise it violates SSP.

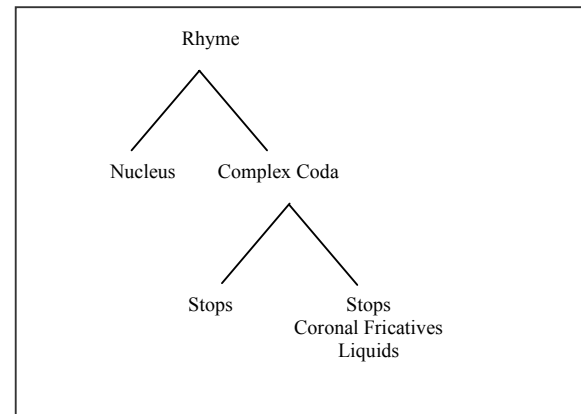


Figure 1 Rhyme Structure for Stops at First Coda Position

The nasal stop [n] tends to come at first position in both word medial and word end CC cluster. It is then followed by a stop (aspirated or un-aspirated), a voiceless fricative or an affricate (aspirated or un-aspirated), as illustrated in Figure 2, e.g. in the words [p nd r h], [m s] and [m nd n].

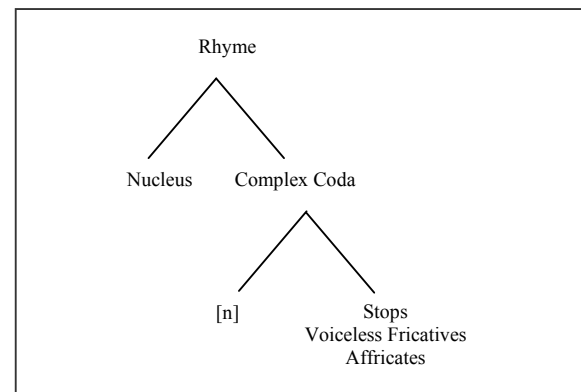


Figure 2 Rhyme Structure for [n] at First Coda Position

The fricative [h] is mostly found at first position in CC cluster and is followed by a nasal, a liquid, a trill, a voiceless fricative or a stop, as illustrated in Figure 3, e.g. the words [m hn d], [d h], [m hz] and [k ht].

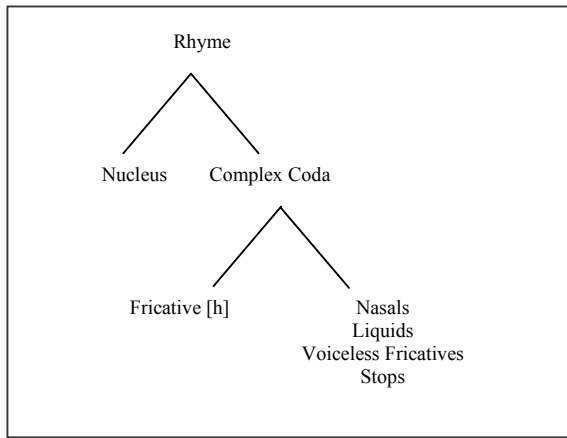


Figure 3 Rhyme Structure for [h] at First Coda Position
Glides neither follow nor are followed by any other consonantal segment except [w] that behaves as fricative in that case. Furthermore, affricates are mostly found at second position in CC cluster.

Liquids are highly sonorant and occur at first position in coda of a syllable. The second consonantal segment can be a stop (nasal or non-nasal), a fricative or an affricate, as illustrated in Figure 4, as in the words [ɾ m ɾɖ], [h ɾf] and [d ɬ].

5.3. Sonority Sequence

Sonority sequence of different classes of consonantal segments (e.g. stops, affricates, glides etc.) is given in the Literature Review Section. One of the purposes of this research is to determine the internal sonority hierarchy of consonantal sounds in each of these classes. As (Goldsmith, 1990) has mentioned a simple sonority hierarchy of consonantal segments for English (pp. 110-112), it should be verified for Urdu language. The sonority of segments is determined through SSP, which might be violated at word boundary. Therefore, a potential syllable, which has complex coda, must occur word-medially to argue for sonority levels of two segments in coda of that syllable.

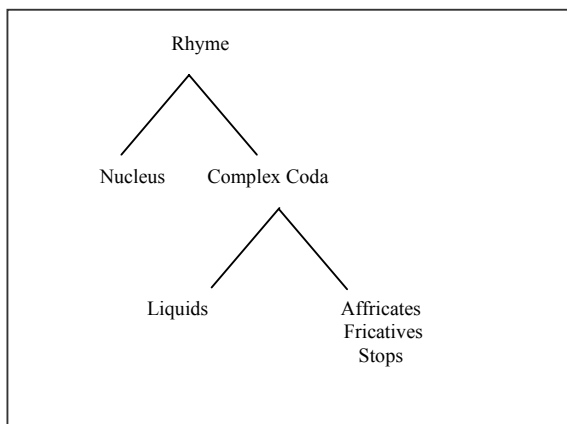


Figure 4 Rhyme Structure for Liquids at First Coda Position

The results reveal that the segments of different classes do occur word-medially as in words [p ɾk .ɪ], [p ɾɖ ɾ], [ɾɖ m ɾɖ], and [d ɔmp .ɪ]; but segments belonging to a same class do not occur together word-medially in coda position. This argues that segments with similar nature have the same level of sonority. Although, the voiced segments are more sonorous than their voiceless versions, but still their difference of sonority level is not enough to allow them to come together word-medially in coda position. For example, voiced affricate [d] is more sonorous than voiceless affricate [t], but still these do not occur together in same consonantal cluster of coda or onset position of a syllable.

6. CONCLUSION

The results reveal that the Urdu has complex coda cluster with at most two consonants. The possible syllable templates for Urdu with complex coda are VCC, CVCC and CVVCC. The syllable template VVCC was not found in Urdu.

Mostly, the first position at CC coda cluster can be a voiceless fricative, nasal, liquid, trill or a labial stop while second consonant is a stop (dental or voiceless velar stop most of the time). If second consonant is a fricative, a trill or a liquid, the first consonant may be a stop, fricative or nasal.

The research has shown that Urdu has a very small set of complex onsets. The segments of different classes occur word-medially in Urdu, but the consonantal segments with similar nature have the same level of sonority. The voiced segments are more sonorous than their voiceless versions, but their difference of sonority level is not enough to allow them to come together word-medially in coda position.

REFERENCES

- Akram, Bilal. "Analysis of Urdu Syllabification Using Maximal Onset Principal and Sonority Sequence Principal". In *Ikhbar-e-Urdu* April May 2002.
- Bokhari, S. *Phonology of Urdu Language* Royal Book Company, Karachi, Pakistan 1985.
- Gazali, Muhammad Ahmad. "Urdu Syllable Templates". in *Ikhbar-e-Urdu* April May 2002.
- Goldsmith, A. *Auto segmental and Metrical Phonology Massachusetts* Basil Blackwell LTD 1990.
- Hussain, Sarmad. *Phonetic Correlates of Lexical Stress In Urdu*. Unpublished Ph.D. dissertation, Northwestern University, IL, USA, 1997.
- Khalid, Nauman. "Existence of [v] and [w] in Urdu Language". In *Ikhbar-e-Urdu* April May 2002.
- Merg, Mirza Ahmad Ali. *Standard Twentieth Century Dictionary* Educational Publishing House, Lal Kuan, Delhi-110006, India.
- Nayyar, Shanza. "Syllable Stress in Urdu" *Ikhbar- e-Urdu* April May 2002.
- Nazar, Muhammad Noman. "Syllable Templates in Urdu Language". In *Ikhbar-e-Urdu* April May 2002.
- Ferozsons Urdu – English Dictionary*, Ferozsons (Pvt.) Ltd., Lahore, 1967.
- Raza. "Urdu Consonantal and Vocalic Sounds". In *Ikhbar-e-Urdu* April May 2002.
- Saleem, Abdul Mannan, Kabir, Hasan, Rafique, Muhammad Mustafa, Khalid, Nauman and Shahid, Syed

APPENDIX

Urdu Words	Complex Coda Clusters
Stop-Stop	
r bt	bt
bd	bd
s bk	bk
pt	pt
k tb	tb
f tk	tk
n kb	kb
w kt	kt
n kd	kd
Stop-Fricative	
h bs	bs
s bz	bz
h b	b
l tf	tf
k ds	ds
m dh	dh
w kf	kf
r ks	ks
n kz	kz
n k	k
Stop-Liquid/Trill	
k bl	bl
d br	br
r tl	tl
f tr	tr
dl	dl
k dr	dr
b kl	kl
f kr	kr
Stop-Nasal	

x tn	tn
r km	km
bn	bn
m tn	tn
Fricative-Stop	
d ft	ft
w fd	fd
w fk	fk
n sb	sb
sp	sp
s st	st
k sd	sd
f sk	sk
d zb	zb
d zd	zd
r zk	zk
j b	b
b h t	t
r d	d
r k	k
s xt	xt
w hb	hb
k ht	ht
m hd	hd
Fricative- Fricative	
h fz	fz
k f	f
n sf	sf
m sx	sx
m sh	sh
h zf	zf
k f	f
xs	xs
xz	xz
b x	x
m z	z
k hf	hf
b hs	hs
m hz	hz
f h	h
Fricative-Glide	
l w	w
m hw	hw
Fricative-Liquid/Trill	
k fl	fl
k fr	fr
m sl	sl
k sr	sr
f zl	zl
d zr	zr
h r	r
b xl	xl
d hl	hl
Fricative-Nasal	
d fn	fn
d sn	sn
h sn	sn
b zm	zm
zn	zn
t m	m
d n	n
r m	m

w hm	hm
m hn̩di	hn̩
m h .	h
Affricate-Fricative	
n d s	d s
d z	d z
Affricate-Liquid/Trill	
f d r	d r
Affricate-Nasal	
h d m	d m
Liquid/Trill-Stop	
s lb	lb
x lt	lt
d ht	ht
l	l
x lk	lk
k rb	rb
f rt	rt
s rd	rd
b r	r
t rk	rk
b r	r
Liquid/Trill-Fricative	
k lf	lf
f ls	ls
t lk	lk
m lh	lh
h rf	rf
rs	rs
t rz	rz
t r	r
s rx	rx
m r	r
Liquid/Trill-Glide	
d lw	lw
Liquid/Trill-Affricate	
m rt	rt

rd m nd	rd
Liquid/Trill-Nasal	
z ln	ln
d m	m
k m	m
Nasal-Stop	
k mb	mb
d omp.i	mp
smt	mt
nt	nt
p nt	nt
x nd	nd
p nd rh	nd
n	n
n	n
k n	n
d nkn	nk
p nk .i	nk
Nasal-Fricative	
ms	ms
rmz	mz
snf	nf
b nsri	ns
tnz	nz
b nh	nh
Nasal-Liquid/Trill	
r ml	ml
mr	mr
Nasal-Affricate	
d nt n	nt
s.f nd	nd
m nd n	nd
Nasal-Nasal	
mn	mn