

URDU SYLLABLE TEMPLATES

MUHAMMAD AHMAD GHAZALI

ABSTRACT

This paper aims at determining the syllable templates of Urdu. A detailed experiment was carried out to analyze five thousand words of Urdu and determine all possible syllable templates. Eleven syllable templates were identified for Urdu and their frequencies of occurrence were determined. The syllable templates that were identified are: CV, CVC, CVCC, CVV, CVVC, CVVCC, V, VC, VCC, VV, and VVC. Later on, it was realized that Urdu has basically only first 6 templates and other 5 are derived from these fundamental templates.

1. INTRODUCTION

Urdu is the national language of Pakistan and is widely spoken in the Indian subcontinent (Pakistan and India), as a first or a second language.

The syllable structure has received quite a lot of attention in the phonological literature of the past few decades. It has been shown that the information about syllable structure of a language is essential for comprehending phonological formation of that language. In the present study the syllable structure of Urdu is studied, showing the nature of the permissible syllable templates in Urdu along with their frequencies of occurrence. Different restrictions on the use of these templates are also discussed.

2. LITERATURE REVIEW & PROBLEM STATEMENT

When talking about syllable templates of a language, the first question that arises is that 'what really a syllable is'? Therefore, it seems rational to start by highlighting the notion of syllable in modern phonology.

2.1 Syllable

Syllable is an essential concept for understanding phonological structure

(Kenstowicz, 1994, p. 250). Trask points out that attempts have been made to define the syllable in terms of muscular contraction and in terms of peaks of sonority but no completely satisfactory definition has been found (Trask, 1997, p. 214). Ladefoged defines the syllable as the smallest possible unit of speech (Ladefoged, 1993, p. 248). He states that every utterance must contain at least one syllable.

The syllable is believed to consist of an obligatory *nucleus* preceded by an optional consonantal *onset* and followed by an optional consonantal *coda* (Kenstowicz, 1994, p. 252). The nucleus and the coda form a unit together, which is called the *rhyme*. Thus a syllable structure can be manifested as shown in Figure 1.

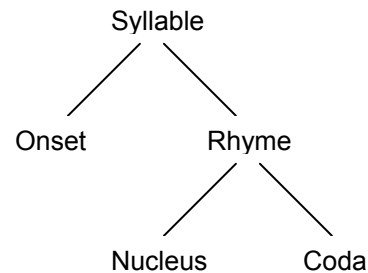


FIGURE 1: STRUCTURE OF SYLLABLE

In all languages restrictions have been found on how many segments can appear in the three positions of the syllable (Goldsmith, 1990, p. 108).

A syllable whose rhyme contains only one element is called a Light syllable and if it contains more than one element then it is called Heavy (Napoli, 1996, p. 81). This concept of heaviness is sometimes conveniently represented by moraic notation. A *mora* is a time unit equivalent to a single vowel or a coda consonant. Thus, a long vowel or a vowel-consonant sequence can be represented in an abstract sense as bi-moraic, or twice as long as the short vowel. Defining this unit is useful because it simplifies the analysis of some stress languages (Hussain, 1997, p. 11).

2.2 Syllable Template & Syllabification Principles

Hogg & McCully define a template or a syllable template as an abstract tree structure onto which all syllables would have to fit in order to be recognized as acceptable syllables in a particular language (Hogg & McCully, 1987, p. 41). There may be several restrictions and permissions, which contract or expand upon the basic template(s) in a language. The process of associating a linear string of segments with a syllable structure or template is called *syllabification*.

CV is considered to be the most common and basic type of syllable (Napoli, 1996, p. 81). It exists in almost all languages (Kenstowicz, 1994, p. 253) and other types of syllables can be formed by changing this basic template. These templates can contain complex onsets or codas. The construction of these complex onsets and codas follow a sonority pattern that is referred to as the 'Sonority Sequencing Principle' (SSP). The sonority of a sound is its loudness relative to that of other sounds with the same length, stress and pitch (Ladefoged, 1993, p. 245). The SSP requires the onsets to rise in sonority towards the nucleus and codas to fall in sonority from the nucleus (Kenstowicz, 1994, p. 254). The sonority hierarchy that is followed by this principle is listed in (1) below (Goldsmith, 1990, p. 111). Note that the sonority decreases from the top to bottom and within each category voiced segments are more sonorous than the voiceless ones.

- (1) Vowels
 Low vowels
 Mid vowels
 High vowels
 Glides
 Liquids
 Nasals
 Obstruents
 Fricatives
 Affricates
 Stops

Goldsmith states that the SSP is intended as a necessary condition for basic syllabification and not as a universal statement of syllables possible in any language. It means that there can be

languages whose syllable templates violate the sonority sequencing principle (Goldsmith, 1990, p. 111).

Another principle used for syllabification is the 'Maximal Onset Principle' (MOP). This principle states that when there is a choice as to where to place a consonant it is placed in the onset rather than the coda.

2.3 Urdu Syllable Structure

Urdu belongs to the family of New Indo-Aryan (NIA) languages, which is a sub-branch of the Indo-European languages (Hussain, 1997, p. 39). It is the national language of Pakistan and is spoken by at least 104 million people (including second language speakers) in more than 20 countries (Ethnologue, 2001). It is quite intelligible with Hindi but has formal vocabulary derived from Arabic and Persian. Urdu writing system uses Arabic script in Nastaliq style with several extra characters.

Much research has not been done on the phonology of Urdu and little material is available about the syllable structure of Urdu. Only Bokhari (1985) and Hussain (1997) have done some work on the syllable templates of Urdu. Bokhari has listed the syllable templates of Urdu (Bokhari, 1985, p.18) but these templates are incomplete and ambiguous. The description and discussion of these templates is also quite confusing. Also, he has not listed the frequencies with which these templates occur.

Hussain has listed, based on his pilot work, some templates in his Ph.D. dissertation (Hussain, 1997, p. 42) but he has himself written in a footnote that a detailed study is needed to confirm these templates and more research needs to be done to determine all the constraints on syllable construction in Urdu.

The reviewed literature indicates that the syllable structure or the syllable templates of a language provide a better understanding of the phonological properties of that language. As seen from the above discussion on the previous work done on syllable patterns of Urdu, it becomes clearer that a detailed study is needed to determine

and examine the permissible syllable templates of Urdu along with their relative frequencies. The following work is devoted to this investigation.

3. METHODOLOGY

The first step in the process of identification of syllable templates of a particular language is to analyze the words of that language to determine the permissible syllable patterns. The selection of these words is an important task. Words chosen should be ample in quantity, so that the less frequent templates are also revealed. These words should be chosen from a reliable dictionary.

The dictionary that will be used to select words in this particular experiment is *Feroz-ul-Lughat Urdu Jaibi* by *Ferozsons*. Five thousand (5000) words will be chosen from this dictionary for the purpose of analysis. The criteria for selection of words will be the frequency of occurrence of the first letter of the word in the dictionary. For this purpose letters will be sorted according to the No. of pages occupied in dictionary by the words starting with that letter. After this list of letters is compiled and arranged properly, the next step is the selection of letters from this list. The letters will be chosen in such a way so that the most frequent letter is analyzed first and then the least frequent one is analyzed. This process of alternating selection of words, firstly the most frequent letter and then the least frequent, will continue until the analysis of all the 5000 words is completed.

The above-mentioned scheme of letter selection is adopted because the less frequent letters may present some special restrictions on the syllable templates. If words starting with a particular letter are less, then it means that the letter has some characteristics that restrict the formation of legal words. Analysis of such letters can be very helpful in finding out some exceptional templates or exposing some restrictions on the use of existing templates. On the other hand, the analysis of more frequent letters will help in finding out the frequency of occurrence of various templates. So, it necessitates that the analysis of a more-frequent letter should be followed by the analysis of a less-frequent letter. This

procedure should continue until the required No. of words are analyzed.

The analysis of a word will consist of writing down the word in Urdu and then writing the syllable template(s) of that word, while indicating the syllable boundaries. As an example, analysis of two words is shown below in (2):

(2)

/sa.hɪl/ ساحل (Coast)
CVV.CVC

/nɪg.ran/ نگران (Supervisor)
CVC.CVVC

The transcriptions shown in these examples are not required and have only been shown to demonstrate that long vowels are represented as VV and short vowels as only a single V. This is because a short vowel in Urdu is mono-moraic and a long vowel is bi-moraic (Hussain, 1997, p. 44). Urdu has seven long vowels and three short vowels (Hussain, 1997, p. 148). In the present experiment the long vowels (i, e, æ, u, o, ɔ, a) have been represented as VV and the short vowels (ɪ, ə, ʊ) as V.

After the analysis of words, the next thing to do is to determine the frequency of occurrence for each template. For this purpose, a frequency distribution of the syllable templates would need to be made. In this frequency distribution, the No. of occurrences of each template will be noted down. The syllable templates of Urdu with their frequencies of occurrence are listed in the Results section.

Once the relative frequencies of templates are determined, the final step is to analyze the identified templates to find restrictions on the use of these templates. These restrictions can be of various kinds e.g. there may be a template that only occurs at the beginning or at the end of a word. Further, some templates can be more frequent in the words starting with a particular letter. There may be more constraints on the consonants or vowels of a syllable template. All of these restrictions

need to be determined in order to fully specify the syllable structure of any language. In the case of Urdu, these restrictions are discussed in the Discussion section.

4. RESULTS

According to the above-mentioned methodology, the analysis of five thousand words of Urdu was done. The analysis revealed that Urdu has eleven syllable templates. These templates with their relative frequencies of occurrence (expressed as percentage) are listed in Table 1.

TABLE 1 URDU SYLLABLE TEMPLATES WITH THEIR FREQUENCIES OF OCCURRENCE

Sr. No.	Template	Frequency (percentage)
1	CV	16.0
2	CVC	21.8
3	CVCC	3.0
4	CVV	37.0
5	CVVC	16.5
6	CVVCC	0.4
7	V	0.2
8	VC	0.8
9	VCC	0.03
10	VV	2.8
11	VVC	1.5

The letters that were analyzed in this experiment are listed in Appendix A.

Urdu is one of those languages that make a three-way weight distinction between its syllables. In Urdu, mono-moraic syllables are *light*, bi-moraic syllables are *heavy* and the syllables having three or more moras are *super-heavy* (Hussain, 1997, p. 47). The results indicate that Urdu has 2 light syllables (CV, V), 4 heavy syllables (CVC, CVV, VC, VV) and 5 super-heavy syllables (CVCC, CVVC, CVVCC, VCC, VVC). The frequency analysis of these three types of syllables reveals that heavy syllables are the most frequent (62.4%), then comes the super-heavy syllables (21.4%) and light syllables are the least frequent (16.2%).

If we look at the frequencies of various syllable templates, it can be seen that CVV (37%) is more frequent than CVVC (16.5%).

Based on this observation, we may infer that open syllables are more frequent in Urdu, as compared to the closed syllables but this is not the case when we compare the frequencies of the syllables CV (16%) and CVC (21.8%). So, the problem is how to justify the more frequent syllables. Looking at it from the mora point of view this problem is easily solved. As it can be seen that the templates CVV and CVC are more frequent than CVVC and CV, because they contain two moras i.e. they are heavy and the other two syllables (CV and CVVC) are light and super-heavy, respectively. Thus, it can be said that heavy syllables are more frequent in Urdu.

5. DISCUSSION

Following is the analysis of each of the eleven syllable templates identified in the previous section:

5.1 CV

As pointed out earlier, it is the core syllable and exists in almost all the languages. In Urdu, this syllable can occur word initially or medially but not word finally. There is no restriction on which consonants can occur in this template. The examples of this template are listed in (3):

(3) /sɪ. t̪əɪm/ (Oppression)
CV. CVC

/ho. nə. hɑr/ (Intelligent)
CVV.CV.CVVC

5.2 CVC

One restriction that has been found on the occurrence of this template is that the consonants in the onset and coda are allowed to be same only if they belong to the set /t/, /t̪/, /l/, /ʃ/, /s/, /b/, /m/, /p/, /t̪ʃ/. Examples of such words are /l̪əɪ.kɑr/ (Challenge) and /sɪs.ki/ (sobbing).

5.3 CVCC & CVVCC

According to Hussain (Hussain, 1997, p. 42), when there are two coda consonants, the first consonant is limited to the voiceless

fricatives (/f/, /s/, /ʃ/ or /x/) or nasals (/n/ or /m/) and second consonant is restricted to stops. The current experiment revealed that first coda consonant could also be one of the following: /l/, /z/, /r/, /ʁ/, /b/, /t̪/, /k/. However, these phonemes can only occur within the template CVCC and not in CVVCC. The phonemes that can occur in both the templates are /f/, /s/, /ʃ/ or /x/. The nasals (/n/ or /m/), also, can only occur in the template containing a short vowel (i.e. CVCC). Some examples of the templates CVCC and CVVCC are given in (4) and (5), respectively:

- (4) /səxt̪/ (Hard)
CVCC
- /kɪzb/ (Lie)
CVCC
- (5) /kof̪t̪/ (Anguish)
CVVCC
- /jɑd̪. d̪ɑʃt̪/ (Memorandum)
CVVC.CVVCC
- /bər. xɑst̪/ (Dismissal)
CVC.CVVCC

Goldsmith (Goldsmith, 1990, p. 108) expresses that some languages explicitly allow word-initial or word-final elements to remain extrasyllabic i.e. this extra segmental material, at the end or beginning, could not be syllabified according to the principles that appear to hold word-internally. As seen from the examples given in (5), it appears that the template CVVCC occurs at the end of words. Thus, it can be said that the last consonant in this template is extra-syllabic and has been, extra-syllabically, attached to the template CVVC to produce the template CVVCC.

5.4 CVV

This is the template that occurs most frequently (37%) in Urdu. No restriction has been found on the occurrence of this template i.e. it can occur word initially, medially or finally. Also, there is no constraint on what consonant may be in the

onset position. Few examples, of the use of this template, are given in (6):

- (6) /t̪ɑ. kəʃt̪. vər/ (Strong)
CVV.CVVC.CVC
- /t̪ɪ. d̪zɑ. rəʃt̪/ (Trade)
CV.CVV.CVC

A question may arise that CVC is more of a complete template as compared to CVV, then why the template CVV is more frequent than CVC. The reason behind this is that there is less effort required in saying CVV as compared to CVC, because of not saying an extra consonant and just lengthening the vowel. This point is further supported by the fact that there are several words in Urdu that end in ة e.g. سادہ (simple), instead of being spoken as /sɑ. d̪əh/ it is actually spoken as /sɑ. d̪ɑ/. What is happening here is that the short vowel /ə/ is being lengthened and changing into /ɑ/ and /h/ is being deleted. To illustrate this point further, some more example words are given in (7):

- (7) /nuk. t̪ɑ/ (Point)
/kəʃt̪. rɑ/ (Drop)
/t̪ər. kɑ/ (Legacy)

Thus, it can be said that such words bring a decrease in the frequency of the template CVC and increase the frequency of CVV.

5.5 CVVC

There is no restriction on the occurrence of this template and it can occur word initially, medially or finally. However, word medially its use is less frequent. This may be because of the fact that Maximal Onset Principle (MOP) would have been applied on this template, to remove the consonant from the coda and put it into the onset of the following syllable. Thus, producing a CVV template instead of CVVC and this may be another factor behind the more frequent use of the CVV template. To illustrate the use of this template, few examples have been shown in (8):

- (8) /bɑɣ. bɑn/ (Gardener)
CVVC.CVVC

/t̪əs.vir/ (Picture)
CVC.CVVC

5.6 V

This is one of the templates that are very less frequent. This fact is understandable because such templates are unusual and not very common in the languages of the world (Kenstowicz, 1994, p. 254). This template can occur at the beginning or middle of a word but has not been found to occur at the end of a word. Examples of this template are given in (9):

(9) /ɪ.ləm/ (Knowledge)
V.CVC

/bɪl.ʊ.mum/ (Generally)
CVC.V.CVVC

5.7 VC

This is also a less frequent template and it can occur word initially, medially or finally. Some example words containing this template are given in (10):

(10) /ɪn.sɑf/ (Justice)
VC.CVVC

/sɑ.ət̪/ (A moment)
CVV.VC

5.8 VCC

This is the least frequent template in Urdu and has only been found to occur 4 times in 5000 words. Two of these words are listed in (11):

(11) /fɑ.t̪ɪ.rʊl.əkɪ/ (Unintelligent Person)
CVV.CV.CVC.VCC

/bər.əks/ (On the contrary)
CVC.VCC

5.9 VV

This template can occur initially or medially but not word finally. About 80% of the time this template has occurred in the words that

start with the letter $\bar{\text{f}}$ (pronounced as /ɑ/). Few examples are given in (12):

(12) /ɑ.t̪ɪʃ/ (Fire)
VV.CVC

/sə.ɑ.d̪ət̪/ (Auspiciousness)
CV.VV.CVC

5.10 VVC

This template can occur word initially, medially or finally. A vast majority (about 70%) of the occurrences of this template have been in the words starting with the letter $\bar{\text{f}}$. Some examples are given in (13):

(13) /ɑn/ (Dignity)
VVC

/nəs.buɪ.æn/ (Ideal)
CVC.CVC.VVC

In the examples given for the onset-less syllables, there is a glottal stop (ʔ) before the first vowel of each onset-less template. This stop is derived from the Arabic background of Urdu and is not pronounced in Urdu, and because of this the onset-less syllables seem to occur in Urdu. So, it means that the templates starting with a vowel are not really syllable templates of Urdu. Therefore, it seems plausible that the syllable templates without the onset (V, VC, VCC, VV and VVC) have, underlyingly, a consonant in the onset. So, the appropriate representations of these templates are CV, CVC, CVCC, CVV and CVVC, respectively. Thus, it can be said that Urdu has six basic syllable templates and other five are derived from these basic templates. As it can be seen from these templates that Urdu does not allow complex onsets (onsets having more than one consonant) to occur, however, complex codas are allowed.

6. CONCLUSION

There is no doubt that the syllable is an obligatory unit of phonological organization of a particular language and thus a principle research objective. Phonologists are interested in the structure of the syllable, since there appear to be interesting

observations to be made about which phonemes may occur at the beginning, in the middle and at the end of syllables.

This study focused on the syllable structure of Urdu, after giving a brief introduction to the various terminologies used in modern phonology regarding syllables. Eleven different syllable templates were identified for Urdu but later analysis revealed that underlyingly Urdu has only six fundamental syllable templates (CV, CVC, CVV, CVCC, CVVC and CVVCC). Frequency analysis of these underlying syllable templates showed that the template CVV is most frequent (37%) and the template CVVCC occurs least frequently (0.4%).

7. REFERENCES

Bokhari, S. 1985. *Phonology of Urdu Language*. Royal Book Company, Karachi.

Ethnologue: Languages of the World, 14th Edition, 2001 © Summer Institute of Linguistics, URL: <http://www.sil.org>

Feroz-ul-Lughat Urdu Jaibi, Ferozsons, Lahore.

Goldsmith, J. 1990. *Autosegmental and Metrical Phonology*. Blackwell, Oxford.

Hogg, R. and McCully, C. 1987. *Metrical Phonology: A Coursebook*. Cambridge University Press, Cambridge.

Hussain, S. 1997. *Phonetic Correlates of Lexical Stress in Urdu*. Unpublished Ph.D. dissertation, Northwestern University, IL, USA.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Blackwell, Cambridge.

Ladefoged, P. 1993. *A Course in Phonetics*. Harcourt Brace Jovanovich, New York.

Napoli, D. 1996. *Linguistics: An Introduction*. Oxford University Press, New York.

Trask, R. 1997. *A Student's Dictionary of Language and Linguistics*. Arnold, London.

8. APPENDIX A

Following are the letters that were analyzed in this experiment, along with the number of words that were examined for that particular letter:

TABLE A.1 LIST OF ANALYZED LETTERS

Letter	No. of words analyzed	Letter	No. of words analyzed
ث	761	ف	220
ب	680	ل	173
س	675	ہ	136
آ	408	ط	124
ش	354	ٹ	120
ن	336	ی / اے	73
ر	314	ؤ	57
ک	305	ثر	8
ق	256		