Rafia Bokhari, Sadaf Pervez

Syllabification & Resyllabification in Urdu

Abstract: *This paper tends to describe a n approach to syllabification and resyllabification in Urdu language. It focuses on the applicability of different syllabification algorithms. Different scenarios under which resyllabification is required are distinguished and the process of resyllabification is explained.*

Keyword: *Syllabification, Resyllabification, Syllable templates, Urdu.*

1. INTRODUCTION

Syllabification of a language is important for phonological analysis. Syllable is a phonological entity that is necessary to explain the existence of some other important phonological entities like stress etc.

In the scope of this paper we will follow the following conventions:

The syllable is represented by its symbol σ.
C represents a single consonant.
C* donates zero or more consonants.
V represents short vowel, taking only one time slot.
VV represents a long vowel with two time slots.

Optional consonants and vowels are represented in parentheses (C) and (V) respectively.

2. LITERATURE REVIEW

Before going in to the details of syllabification and re-syllabification rules, we first need to get familiarized with some important related terms.

2.1 Syllable

Laderfoged defines syllable as the smallest possible unit of speech (Laderfoged, 1993, p.248). Laderfoged believes that every utterance must contain at least one syllable so it can be said that syllable is smallest unit of speech. The role of syllables in phonological theory is controversial (Kenstowicz, 1994). Previously it was thought that the notion of syllable has no official recognition. But due to the modern phonological analysis and the tendency of the native speakers of the language to break speech in syllables, it's clear now that syllable is an important phonological entity and without the concept of syllables, the phonological structure can't be understood properly.

The biggest hurdle in making syllable a definite phonological entity is the difficulty to assign it a comprehensive definition. However, there have been attempts to formulate a comprehensive definition for the syllable. There are also many theories proposed for this like the chest pulse theory (Trask, 1997) however, none of them has proved adequate enough.

As in the example below, the Urdu word /at□bazi/ has four syllables.

/a .t□ .ba .zi/

2.2 Syllable Templates of Urdu

There are different syllable templates in Urdu language. A syllable template can occur word initially, medially or word finally. However, there are certain limitations on these templates like some of these templates are not permissible at the word boundaries while the other are not permissible word medially. The permissible syllable templates of Urdu are given below:

Table 1 Possible Syllable Templates in Urdu

Syllable Template	Allowed Word initially	Allowed Word medially	Allowed Word finally
CVC	A	A	A
CVV	A	A	A
CVCC	A	A	A
CVVC	A	A	A
CVVCC	NA	NA	A
CV	A	A	NA
V	Derived Templates		
VC			
VV			
VCC			
VVC			

A: Allowed
NA: Not Allowed

Later on it was realized that Urdu has basically only 6 templates i.e. CV, CVC, CVCC, CVV, CVVC, CVVCC and other five are derived from these fundamental templates (Muhammad, 2002).

2.3 Syllable Structure

Urdu has an interesting syllabic structure. The syllable has the following tree like structure:

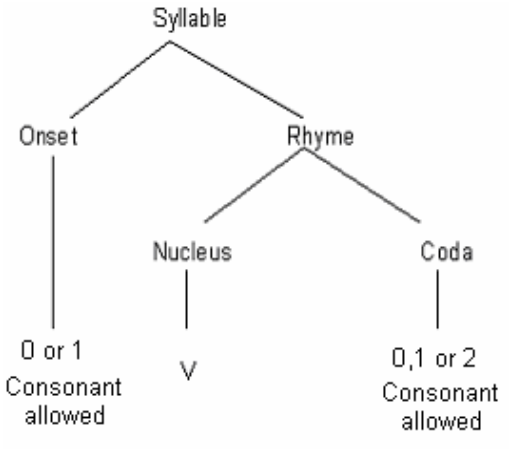

Figure 1 Syllable Structure

For onset maximum one consonant can exist and for coda maximum two consonants can exist in Urdu language.

In this tree, the syllable has three important parts: Onset, Nucleus and Coda. Onset contains consonants. In English multiple consonants can come in Onset position. Whether multiple consonants can come in Onset of an Urdu syllable will be examined in our study. For a syllable to exist, nucleus is the minimal part. Without nucleus, a syllable cannot exist. The nucleus is usually vocalic, however, in English non-vocalic nucleus is also found in some cases. For example:

t a . bl.
CV.CV

N denotes non-vocalic nucleus, which is l in the above case. However in the course of our study we found that Urdu nucleus is always vocalic. The coda part of the syllable contains consonants.

2.4 Syllabification Principles

Languages tend to follow two important syllabification principles Sonority Sequence Principle (SSP) and Maximal Onset principle (MOP).

2.4.1 Sonority Sequence Principle (SSP)

The sonority of sound is its loudness relative to that of other sounds of the same length, stress and pitch (Laderfoged, 1993, p.245). Sonority can also be taken roughly as measurement of openness of sounds. The SSP requires the onset to rise in sonority towards the nucleus and codas to fall in sonority from the nucleus (Kenstowicz, 1994). Sonority increases from the beginning till the nucleus and then falls from the nucleus till the end as shown in Fig. 2.

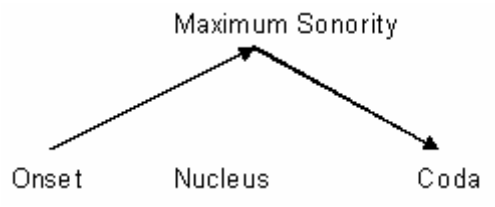

Figure 2 Sonority Sequence Principle (SSP)

This means that the Onset consonants are in increasing order of sonority from left to right and the Coda consonants are in decreasing sonority from left to right. The sonority hierarchy is shown in Fig. 3. The sonority decreases from top to bottom and within each category voiced segments are more sonorant than the voiceless ones (Goldsmith, 1990).

According to Goldsmith SSP is not a universal principle so a language can violate it. But mostly languages follow SSP. As most of the languages, Urdu also follows SSP (Akram, 2002).

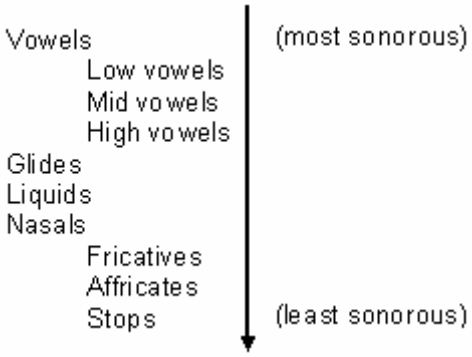

Figure 3 Sonority Hierarchy (from most sonorous to least sonorous)

2.4.2 Maximal Onset Principle (MOP)

In MOP consonants are preferred in the Onset and thus allowing no Coda consonants except for the word final position.

In simple words, MOP means that languages prefer that a floating consonant (a consonant which may occupy either coda or onset) goes to onset position rather than coda position.

2.5 Re-syllabification

Re-syllabification is done due to application of insertion or deletion rules. These rules tend to disturb the syllable template and to restore the permissible templates, re-syllabification is done i.e. the syllabification algorithm is reapplied to get the desired applicable syllable templates.

3. METHODOLOGY

3.1 Subject

In order to do the analysis of syllabification and re-syllabification rules in Urdu, data of 1000 words was collected. For this method first the words were not consulted from the dictionary instead native speakers of

Urdu language were employed and asked to tell the number of syllables in the words in order to have a natural syllabification of these words. Then the words were syllabified according to the syllabification algorithms of Urdu as proposed in this paper. After the application of epenthesis and deletion rules re-syllabification was done.

3.2 Procedure

Three speakers for natural data collection were females. All the speakers were given a list of words and were asked to pronounce it naturally and to divide it in syllables. Then number of syllables in that natural pronunciation was counted and then all the samples were syllabified. After that the deletion and insertion rules were applied to same set of data and then re-syllabification of these samples were done. For this first the dictionary was consulted for having all the consonants and vowels of the word, after that when the deletion or insertion rules were applied, the speakers were asked to retell that how many syllables are there in the word. After that re-syllabification, wherever applicable, was done with the help of our proposed algorithm.

4. RESULTS

4.1 Syllable structure in Urdu

The various configurations of syllables at onset, nucleus and coda are discussed next.

4.1.1 Onset

In Urdu, there can be at the most only one consonant in the onset position. Although there are some cases found in which two consonants appear to come in the onset position but then it was later found that these consonants are extra syllabic. In our research we found that extra syllabic material is only found on word boundary. There are no multiple consonants in the onset position in Urdu syllables.

4.1.2 Coda

There can be at the most three consonants in the coda position of Urdu syllables (Bokhari, 1985). It was also reveled in our research that at the most there could be two consonants in the coda position of Urdu syllables. There are constraints on the nature of these consonants; the major one is SSP; however they are not that strict as pointed out by Hussain. Hussain says when there are two consonants in coda the first is limited to voiceless fricative and the second consonant is limited to a stop (Hussain, 1997). However in the course of our study we found some additional examples e.g. /fərz/, /qərz/, /lərz/, /mərz/ etc. all are not ending in stops. In examples of /wəqt/, /ləxt/, /səkt/ we have first consonant as palatal (k, q), and x. However, note that in these examples the consonant clusters appear in the word boundaries. Our research indicated that the word medial clusters always follow Hussain preposition. So the consonant clusters apparently going against Hussain preposition could be extra syllabic.

4.1.3 Nucleus

In English, there can be non-vocalic nucleus but in Urdu the nucleus is limited to a vowel only. The vowel can be short as well as it can be long. Diphthongs can also come in syllabic nucleus. For example:

k□i, □□i

4.2 Algorithms of Syllabification

Syllabification in Urdu is being done with the help of two algorithms as proposed in the discussion section. It was revealed that the three step algorithm for syllabification of Urdu words is much better than the template matching algorithm because it always shows the correct syllabification however, the template matching algorithm fails from left to right in Urdu as shown in the example in section 5.1.2.

4.3 Re-syllabification in Urdu

In Urdu, re-syllabification is needed at the cases where deletion or insertion rules result in the violation of syllabification principles of Urdu and the number of syllables is changed. For re-syllabification also, the three-step algorithm is employed.

In the case of addition and deletion of vowels normally, the added or deleted number of syllables is equal to the added or deleted number of vowels in that word.

x ə . v a . dz a Trisyllabic

After deletion of əv (Nawaz, 2002)

x a . dz a Bisyllabic

5. DISCUSSION

5.1 Syllabification in Urdu

In Urdu, syllabification can be done with the help of two major algorithms, the three-step algorithm and template matching. Here both of the algorithms are stated below:

5.1.1 Three Step Algorithm

This is a very simple and intuitive algorithm that works as follows:

1. Pick out the vowels from the transcription of the word to be syllabified.
2. Attach one onset consonant to each vowel if possible
3. Place remaining consonants in coda position.

According to this algorithm only one consonant can go in to the onset position of the syllable as this

algorithm works on the constraint that Urdu allows only one consonant in the onset position. However, multiple consonants can come in the coda position of the syllable.

5.1.2 Template Matching

The template-matching algorithm (generic for all languages) says to take out all the possible templates of Urdu language as shown in Table 1 and match it to the transcription of the word from left or right.

The following Example demonstrates this algorithm for Urdu from right to left as well as from left to right. A constraint of this algorithm is that it does not work from left to right in Urdu. This is also demonstrated by this example. It gives wrong syllables when working from left to right.

ərdz.əm.ənd Template Matching
VCC.VC.VCC (Left To Right)

ər.dzə.mənd Template Matching
VC.CV.CVCC (Right To Left)

5.2 Re-syllabification Scenarios in Urdu

5.2.1 Deletion Rules

5.2.1.1 Deletion of Glottal Stop ▫ and re-syllabification rules

Most of the native speakers of Urdu tend to delete the glottal stop in their natural speech. Whenever the glottal stop occurs in the beginning of the word mostly the number of syllable is not changed, only the onset consonant which is a glottal stop in this case gets deleted, so in this case we have a syllable without the onset consonant e.g. after deletion of initial ▫, o.rat (woman) has two syllables and in the first syllable is an open syllable with no onset. Thus only the syllable template has changed from CV to V and the number of syllables remains the same.

Similarly when the glottal stop occurs in the middle or the end of the word, then also the number of syllables are not changed only the syllable template changes favoring to place one consonant in the onset of the syllable.

Some example of deletion of glottal stop and re-syllabification is given in Table 2 (Nawaz, 2002) and then there re-syllabified results are indicated.

Table 2 Resyllabification after deletion of Glotal Stop

Urdu Word	Syllabification	Deletion Rule	Resyllabification	Meaning
/b▫d▫ə̃t▫/	b▫d▫▫ə̃t	[b▫də̃t▫]	b▫.də̃t▫	Addition in Islam
/bə̃r▫ə̃ks/	bə̃r.▫ə̃ks	[bə̃rə̃ks]	bə̃.rə̃ks	Apart from
/tə▫av▫n/	tə.▫a.v▫n	[tav▫n]	ta.v▫n	Cooperation

5.2.1.2 Deletion of Vowels and resyllabification

Re-syllabification is initiated due to the deletion of a vowel in the word. As one important clue of syllabification in Urdu is that the number of syllables in a word is equal to the number of vowels in that word. So, it is obvious that the deletion of a vowel will result in re-syllabification of the word. Some example of deletion of ə in xə context as given in (Nawaz, 2002) are resyllabified after deletion in the following data set

Table 3 Re-syllabification after deletion of ə

Syllabification	Deletion Rule	Resyllabification
/x▫.vab/	[xab]	[xab]
	[xoab]	[xo.ab]
/x▫.va.d▫a/	[xad▫a]	[xa.d▫a]
	[xad▫a]	[xa.d▫a]
/x▫.var/	[xar]	[xar]
	[xoar]	[xo.ar]
/x▫.vast.▫ar/	[xast▫ar]	[xast▫▫ar]
	[xast▫ar]	[xast▫▫ar]
/x▫.van.da/	[xanda]	[xan▫da]
/x▫.vah/	[xah]	[xah]
	[xoah]	[xo.ah]
/x▫.va.hir/	[xahir]	[xa.hir]
	[xoahir]	[xo.a.hir]
/x▫.va.hi▫/	[xahi▫]	[xa.hi▫]
	[xoahi▫]	[xo.a.hi▫]

5.2.2 Epenthesis Rules

In the course of our research we came across a very few epenthesis examples in Urdu. However, we found out that Urdu inserts vowels wherever possible, in order to avoid multiple consonants in the onset position. For example in Urdu word (tree) is normally syllabifies by the native speakers in two ways /də.rəxt/ containing two syllables and /drəxt/, containing only one syllable. /də.rəxt/ is syllabified as under

Figure 4 Syllabification of /d ərəxt/

According to this syllabification, there is only one consonant in the onset position. However /drəxt/ is syllabified as:

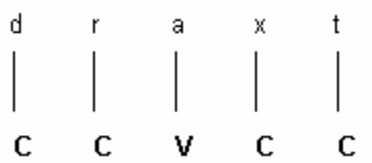

Figure 5 Syllabification of /drəxt/

However /drəxt/ is not a preferred pronunciation in Urdu. Hence it can be concluded that the native speakers of Urdu tend to insert a vowel a between d and r to make the pronunciation natural as in first case.

The epenthesis rules are applied at cases to avoid multiple consonants in the onset position. It can also be proposed that same goes for the consonants in the coda position. Maybe this is the reason why Urdu does not allow more than two consonants in the coda position.

6. CONCLUSION

Syllable is an important phonological entity. The syllabification algorithms of Urdu language strictly follow the SSP in all cases. However, in some cases the extrametricality may come at syllable boundaries. The syllable structure in Urdu is rather a simple one, with one onset consonant and at the most two coda consonants. Urdu always takes a vocalic nucleus. The concept of re-syllabification comes in those cases of Urdu where a vowel gets inserted or deleted.

REFERENCES

-Akram, B. *Analysis of Urdu Syllabification using Maximum Onset Principle and Sonority Sequence Principle.* Akhbar-i-Urdu, p. 177, Lahore, Pakistan, 2002.

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

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

-Goldsmith, J. *Auto segmental and Metrical Phonology*. Blackwell, Oxford, 1990.

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

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

-Laderfoged. *A Course in Phonetics*. Harcourt Brace Jovanovich, New York, 1993.

-Muhammad. *Urdu Syllable Templates.* Akhbar-i-Urdu, Lahore, p.203, Pakistan, 2002.

-Nawaz, S. *Deletion Rules in Urdu Language.* Akhbar-i-Urdu, p. 233, Lahore, Pakistan, 2002.

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