Urdu Syllable: Templates and Constraints

Mazhar Iqbal Ranjha
Minhaj University, Lahore
mazranjha@gmail.com

Abstract

Urdu has 4 heavy syllables or bi-moraic (CVC, CVV, VC, VV) and 5 super-heavy syllables (CVCC, CVVC, CVVCC, VCC, VVC), "C" denotes consonant, "V" light vowel, "VV" long vowel and "CC" denotes consonant cluster. In Urdu, "CC" is not allowed in the onset position. It only occurs in the coda position. This paper aims to consult one of the authentic Urdu dictionaries and search all templates with "CC" consonant clusters and constraints. It also aims to investigate, on one hand, all possible consonant combinations and on other hand try to investigate the syllabic principles followed in Urdu.

1. Introduction

Urdu is national language of Pakistan. "Urdu is second or third highest spoken language in the world. It is approximately spoken by 591,000,000 people in more than ten countries as first or second language" [7]. It has rich phonology which has not been fully explored yet. Every language has its rules and parameters. Some languages take complex onsets and other complex codas. Some have complicated syllable structure other have easy. Some follow one rule and some other. Some languages allow different combinations of consonant clusters in coda or onset positions. As Persian, Arabic, Hindi, Sanskrit, Portuguese and other local languages contribute a lot in the vocabulary of Urdu so these languages continue to cast their influence in the direction of evolution of Urdu. The syllable templates of Urdu are similar to the templates of these languages.

Working on stress in Urdu, Hussain [2] has touched its other aspects as well. Describing syllable weight, he has listed 12 syllable templates in Urdu. He has divided them into simple and complex onset templates and urges further need of study to confirm the complex ones. Besides him, Ghazali [4] and Nazar [5] have worked on it as well. To some extent they agree on the number of templates but differ in their opinion in phonotactic constraints and consonants clusters in syllables. Akram [6] has also investigated the syllabification and phonotactic constraints observed in Urdu and come with some new ideas. This paper aimed to search such clusters and analyze their behavior. For this purpose, an Urdu dictionary *Feroz ul lughaat* [1] was studied. All the words with consonant clusters were collected. Other phonotactic constraints were also investigated.

2. Literature Review

2.1. Syllable

"Syllable is an essential concept for understanding phonological structure" [11, p. 250]. It is an important unit of language but controversial to be defined. "Different 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" [12, p. 214]. "It is relatively easy for people to count the syllable of a word – much easier than counting the segments" [9, p. 250]. Ladefoged [3] says that although everybody can identify it, nobody can define it. He further states that every utterance must contain at least one syllable [3, p. 230]. Hayes [9] calls it the stressed bearing unit. Perhaps everybody finds syllable comparatively easy to define that is why no serious attention has been paid on defining it. "Every speaker has an intuitive notion of how many syllables each word has. It is less easy for speakers to reflect consciously on the internal structure of a syllable" [13, p. 105].

2.2. Syllable structure

Though the native speakers of any language find it easy to tell how many syllables are present in particular utterance yet it is difficult to give is proper definition that can clarify its phonetic and phonological character. The best way to define a syllable is to talk of its

segments. A speech consists of two segments i.e. consonants and vowels. The universal syllable template accepted by most phonologists is denoted by Latin symbol σ (sigma).
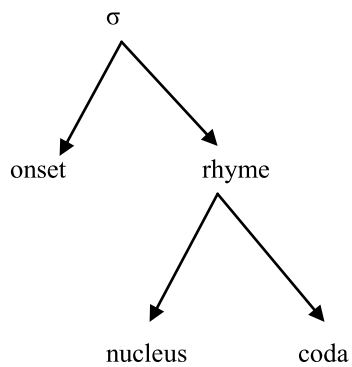

Figure 1

Onset is a consonant or group of consonants that precedes rhyme. Hayes [9] defines onset as consonant or sequence of consonants at the beginning of a syllable. Rhyme of a syllable consists of the vowel and any consonant/s that come(s) after it [3, p. 230]. The nucleus of syllable is the vowel or diphthong found at the syllable's core and functioning as sonority peak. It is an obligatory part for a syllable to have nucleus [9, p. 251]. It is a vocalic part [3, p. 230]. The final part of a syllable, consists of consonant/s, is called coda. The coda and onset are optional parts of a syllable. Onset, nucleus and coda are also called constituents of a syllable.

2.3. Syllabification

An analytical procedure of grouping or dividing a syllable into its components is called syllabification. Bartlett, Kondrak, and Cherry [8] define syllabification as the process of dividing a word into its constituent syllables. Referring Muller et al. and Bouma, Marchand & Damper, they write that technically speaking, syllables are phonological entities that can only be composed of strings of phonemes. Referring Blevins, they say that most of linguists view syllables as an important unit of prosody because many phonological rules and constraints apply within syllables or at syllable boundaries. Citing Goldsmith, Akram [6] names it a process that associates a linear string of segments with a syllable structure. He further writes that from a descriptive point of view, word should be factorable into sequences called syllables, which should have a specifiable internal structure that is roughly constant across the language [14, p. 107].

2.4. General principles of syllabification

"The basis on which syllabification is derived must be (partly) language specific: every language has its own principles of syllabification" [9, p. 251]. Hayes [9, p. 252] describes the following three general principles of syllabification for any language.
- Finding the syllable nucleus
- σ Syllabic affiliation of consonants
- An outline scheme for syllabification

2.5. Theories of syllabification

"There is debate as to the exact structure of a syllable" [8, p. 309]. There is some agreement between the linguistis about a nucleus preceded by onset and followed by coda being constituents of a syllable. A syllable is language specific. Every language has different typological parameters. Number of theories concerning syllabification have been presented by different linguists in different eras.

2.5.1. The Legality Principle. According to this principle "a syllable is not allowed to begin with a consonant cluster that is not found at the beginning of some word, or end with a cluster that is not found at the end of some word" (Goslin and Frauenfelder quoted in [8, p. 309]). Giving an example of an English word *admit,* he says that according to this principle, it is written as [əd-mit]. It cannot be written as [ə-dmit] because no word in English starts with [dm] cluster. In the same way an Urdu word [əlhəmd] cannot be syllabified as [ə-lhəmd] because no cluster is allowed in Urdu in the onset position.

This principle has its limitations and cannot be accepted as universal principle for all languages. Giving an example of a word like *askew* [əskju], [8] comments that "this principle cannot rule out any of [ə-skju], [əs-kju], or [əsk-ju], as all three employ legal onsets and codas".

2.5.2. Maximal Onset Principle (MOP). According to this rule, maximum consonants are preferred in the onset position. "A consonant which may in principle occupy either rhyme or onset will occupy onset position" (Trasj, 1996, p. 217). In Maximal Onset Principle "the consonants are preferred in the onset and thus allowing no coda consonants except for the word final position" (Golsmoth, 1990, p. 128 as in [6]). In a word where there are more than two consonants, leaving the one in the coda of preceding syllable, the rest will go to onset of following syllable For example *escritoire* [eskritwaː] can be syllabified as [e-skrit-waː], [esk-rit-waː] or [es-krit-waː]. According

to MOP, we assign the position of onset to maximum consonants so the division [es-krit-wa:] is correct. And the syllabification of [əskju] will be [əs-kju], Urdu language is very sensitive to onset. It allows only one consonant in the onset position. There are very few words in Urdu with have clusters of more than two consonants. For example an Urdu compound word [ʔərzmɚdʼ] has three consonants[rzm], According to this principle, it will be syllabified as [ʔr.zmɚdʼ], which is not allowed in Urdu. So it can be said that all languages do not follow this principle either.

2.5.3. Maximal Coda Principle (MCP). For syllabification, Maximal Coda principle(MCP) is also used. Opposite to MOP, this principle prefers maximum consonants in the coda allowing no onset consonant except for the word initial position. For example the word [əskju] will be syllabified as [əsk-ju] and [eskritwa:] as [esk-rit-wa:]. As mentioned earlier, Urdu is very sensitive to onsets and it does not follow MCP. In Urdu syllable, if there is only one consonant, then it is preferred in onset. In case of two consonants, the first one will go the onset and the other to coda. If there is cluster of three consonants, very rare case, then it will follow the MCP i.e. two consonants will go to coda position leaving one in the onset position. For example a word [ʔərzmɚdʼ] will be syllabified as [ʔərz-mɚdʼ], This principle also has flaws to be accepted universally.

2.5.4. Sonority Sequence Principle (SSP). This principle states that sounds should occur on their sonority basis. The sonority of sound is determined primarily by the size of their resonance. "The sonority of a sound is its inherent loudness, holding factors like pitch and duration consonant (Crystal quoted in [8, p. 309]). Ladefoged [3, p. 227] defines as "the sonority of a sound is its loudness relative to that of other sounds with the same length, stress and pitch". Citing Selkirk, [8] describes that "SSP states that sonority should increase from the first phoneme of the onset to the syllable's nucleus, and then fall off to the coda". One can observe it by producing vowel and consonant sound alternatively. The vocal tract is more open while producing vowel sound than that of consonant. According to this rule, sonority slope of sounds rises from onset to nucleus and then falls to coda. For example, an English word *vintage* [vintiʤ] cannot be syllabified as [vi-ntiʤ] because [n] is more sonorant than [t]. [t] being stop is least sonorant and according to SSP, sonority should increase from first phoneme to the nucleus i.e. [nt] is not possible cluster. In Urdu

word [ʔərz-mɚdʼ], the coda of first syllable [rz] follows SSP as liquid [r] is more sonorant than fricative [z]. The sonority hierarchy [14, p. 111] is listed in the figure2.

> Vowels
> > Low vowels
> > Mid vowels
> > High vowels
> Glides
> Liquids
> Nasals
> Obstruent
> > Fricatives
> > Affricatives
> > Stops

Figure 2: sonority hierarchy by Goldsmith

According to Goldsmith, this is "necessary condition for basic syllabification and is universally accepted with few exceptions".

2.6. Urdu Language

"Urdu is national language of Pakistan and spoken by more than 100 million people across a more than score countries" [15, p. 01]. "It is popularly regarded as offspring of Persian. It borrows words from different languages to expand its vocabulary. Major languages participating in the camp of Urdu are: Persian, Arabic, Portuguese and English" (Saksena, 1990 quoted in [10]). There are also considerable words of Sanskrit and Hind in Urdui. "Urdu belongs to the family of New Indo-Aryan (NIA) language, which is a sub-branch of the Indo-European language"[2, p. 39]. "Indo-European language family is the most widely studied language as more than half of the world's population speaks one or more of these languages either as a mother tongue or as a business language" [7]. "Urdu and Hindi both belong to NIA Language. They are different literary styles based on the same linguistically defined sub dialect" [16, p. 27]. "In spite of having the same origins and having a very similar linguistic structure, Urdu phonetics and phonology have diverged form Hindi phonetics and phonology. The divergence is perhaps caused by the strong Perso-Arabic influence on Urdu and the strong Sanskrit influence on Hindi" [2, p. 40].

2.7. Urdu Templates

As Persian, Arabic, Hindi, Sanskrit, Portuguese and other local languages contribute a lot in the

vocabulary of Urdu so these languages continue to cast their influence in the direction of evolution of Urdu. The syllable templates of Urdu are similar to the templates of these languages. Not much research has been done on Urdu phonology and little material is available about the structure of Urdu syllables. The work found on Urdu templates is only done by [17] and [2]. As mentioned earlier, the identification of a syllable of a language is intuitive to native speakers but very hard to define. Different theories about the syllabification have been given above but none of these has proved adequate. The syllable is considered as an abstract unit of prosodic organization through which a language expresses its phonology.

Recently, among the United States linguists, there has been found a trend of defining "mora as the element bearing phonological weight". [18] describes that the linguists like Hyman, McCarthy & Prince, Ito, Hayes, Archangeli are of the opinion that "mora plays a major role in syllabic structure". In the theory of Hayes, moras have replaced syllables altogether. The basic concept expressing syllabicity is the Weight Unit (WU). Each segment has a WU.

2.7.1.   Mora. Han in [19] defines mora as a "unit equal to a short syllable". Ladefoged defines it as "a unit of timing". Long vowels are often considered to be bimoric, whereas short ones are monomoraic. Husssain describes it as "a time unite equivalent to a single short vowel". However, a mora is not a "species of sound but rather an elementary prosodic unit …… like the syllable ….intervening between the [syllable] and the phonetic string" (McCarthy and Prince, and Hayes as in [11, p. 392]).

2.7.2. Moraic Structure. Syllables are divided into moras, which function as WU. Moraic structure of languages can vary. For example, in some languages like Latin, both CVV and CVC syllables are counted as heavy (figure 2) and CV as light. Other languages, CVC does not count as heavy (figure 3).

2.7.3.   Moraic Languages. Moraic languages are those in which "the mora plays a part in the phonology or the metrical system". The distinction between mora-timing and syllabic-timing languages is not clear. From terminology "mora-timing" does not mean "moraic". In a mora-timing each mora takes approximately the equal time to be pronounced. It means that a heavy (2-mora) syllable will take twice as long as a light mora. So a moraic language is not essential to be moraic-timing language. These two ideas stand quite apart.

2.7.4.   Syllable template. The study reveals that "the syllable templates of any language provide a better understanding of the phonological properties of that language". The templatic syllabification permits a simpler and more successful analysis of a language. "Syllable templates represent a fixed static set of constraints that dictate the syllable structure in the language concerned". "Templatic syllabification may need some revision" [11, p. 276] but it still stands valid for elementary syllable inventory like Arabic.
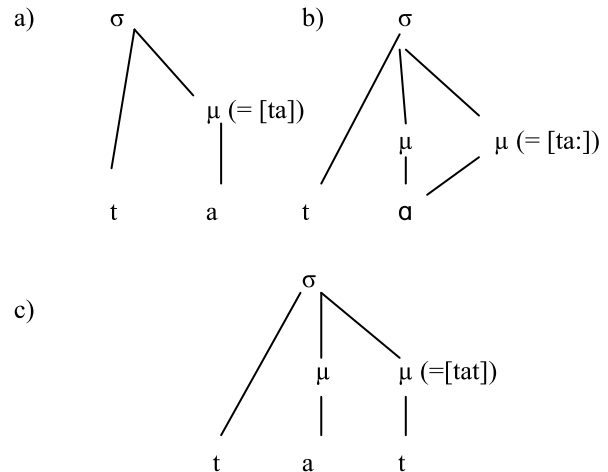
(adapted from www.rnsoke.home.xs4all.nl/CV/publicaties/proefschrift/diss2.pdf)

Figure 3

2.7.5.   Syllable Structure of Urdu Templates. "A syllable template is formulated in terms of sequence of consonants and vowels" [5]. For example, "a syllable template of Arabic language is CV(V)C, where C denotes a consonant, V denotes vowels and ( ) stands for optional templatic element" [11, p.270]. Using the moraic concept, it can be said that a short vowel in Urdu is mono-moraic and long vowel a bi-moraic. In Urdu both vowels and coda consonants are moraic. Consonant clusters in the coda of a syllable are bimoraic. Open syllable with short vowels are mono-moraic. Closed syllable with short vowels and open with long vowels are bi-moraic. "Closed syllables with long vowels or with short vowels and coda cluster are tri-moraic" [2, p. 44-45]. "The moraic differences are represented as a difference of the weight of the syllables. Mono-moraic syllables are called *light*, bi-moraic are called *heavy* and tri-moraic syllables are called *super heavy"* [2, p. 45]. Urdu language counts each short vowel and consonant as mono-moraic, long and nasal vowels bi-moraic and consonant cluster is also bi-moraic. In Urdu two consonants are allowed in the coda position. (See 2.5.3 and 2.5.4).

2.8. Templates and Constraints

Though Urdu is one of the largest spoken and understood language of the world yet little work has been done on its phonetic and phonology. Referring [17, p. 17-19], Akram [6] relates that "a word is made up of at least two sounds a consonant and a long vowel, but no words begin with a long vowel nor with consonants r, rʰ or ŋ nor ends in ŋ. Short vowels ə, cannot occur consecutively within a word nor can any one of them follow the middle consonant of three consonants syllable. The biggest Urdu word is tri-syllabic hence complex words containing more than three syllables are compressed and sounds are assimilated to three syllables".

Talking about phonotactic constraints, [2] has related that open syllables with short vowels do not occur in the word final position. He further says that there can be complex codas and complex onsets in Urdu syllables; however, "there are limitations on formation of these complex onsets and codas" [2, p. 41]. For the formation of syllable, he describes two conditions. "First of all sonority sequencing principle should be satisfied. Secondly, these complex codas can contain at most two consonants. If there are two consonants in the onset, the second consonant in the onset is limited to the glides /w/ or /y/ or may be /h/". Relating the position of two consonants in the codas, he says that "first consonant is limited to a voiceless fricative (/f/, /s/, /ʃ/ or /x/) or nasals (/m/ or /n/)". He is also of the opinion that "alveolar flap cannot occur in the onset position" [2, p. 42].

Ghazali [4] and Nazar [5], in their reports, have described eleven (CV, CVC, CVCC, CVV, CVVC, CVVCC, V, VC, VCC, VV, VVC) syllable templates, they found in their researches.

Ghazali [4] has divided these templates into five categories. He enlists 2 light syllables or mono-moraic (CV, V), 4 heavy syllables or bi-moraic (CVC, CVV, VC, VV) and 5 super-heavy syllables (CVCC, CVVC, CVVCC, VCC, VVC). Both of the above persons agree on CVV to be the most and VC, CVVCC and V least frequent syllable respectively.

Relating the phonotactic constraints, Ghazali [4] claims to have found one restriction that in CVC template, "consonants in the onset and coda are allowed to be same only if they belong to the set /t/, /ʈ/, /l/, /ʃ/, /s/, /b/, /m/, /p/, /ʧ/. Talking about Hussain's [2, p. 42] opinion about phonotactic constraints, he writes that first coda consonant could also be /l/, /z/, /r/, /ɣ/, /b/, /ʈ// or /k/ other than fricatives or nasals". He concludes his discussion about templates by saying

that there are only six basic templates in Urdu and rest five have been derived from these basic ones.

Akram also agrees with Ghazali and Nazar on the number of templates found in Urdu. Discussing the syllable structure of Urdu, he relates that "Urdu is onset loving language as if there is only one consonant in between two vowels then it prefers it in the onset rather than in the coda". For example [azar] will not be pronounced as [az.ar] but [a.zar]. If there are "two consonants together at the end of the word than they both will go in the onset and the coda respectively i.e. first will go in the coda of the first syllable and the second in the onset of the second one". For example [abdoz] will not be as [a.bdoz] or [abd.oz] but [ab.doz]. The cases of three cons active consonants occurring in the middle of the syllable are very rare. In case of their presence, the first two will go to the coda of the first syllable and third in the onset of the next syllable. For example [gondni] will be [gond.ni] not as [gon.dni].

2.9. Summary

Syllable, the smallest unit of speech sound, consists of three constituents; onset, nucleus and coda. There are different syllabification theories, some of which have been discussed in this paper. Every language has its constraints and parameters. Some languages take complex onsets while other complex codas. Urdu is very sensitive to onset position and takes only one consonant. In the coda position it takes maximum two consonants. There are some sounds which are orthographically present but missing from the spoken language. Sometimes speaker use epenthesis to make pronunciation easy. The fundamental templates found in Urdu are CV, CVC, CVV, CVCC, CVVC and CVVCC where C denotes consonant and V denotes vowel. Urdu is moraic language where short vowel is mono-moraic, long and nasal vowels are bi-moraic. Consonant in Urdu is also mono-moraic. With the deletion of some sound or usage of epenthesis, the template undergoes a change. For example, orthographically an Urdu word *abr* [ʔəbr] has CVCC template structure but when a speaker deletes glottal stop, it becomes [əbr] with VCC template. In the same way [sərʰ-ɑ-ne] has CVC-VV-CVV template structure but now in present Urdu it is produced as [sər-hɑ-ne] with CVC-CVV-CVV template structure. With the deletion of [h] sound from [sʊbh] (mornig), it becomes [sʊbɑ]. Hence template CVCC changes to CVCVV. In same way *sang* [səng] (accompany) has CVCC and when [n] nasalizes the preceding vowel, it becomes [sə̃ŋ] with CVVC. Urdu

word [ʔəql] (wisdom) has CVCC template and when [ə] is inserted [ʔəqəl], templates becomes CVCVC.

This paper aims to study the syllables templates, consonant clusters, constraints in Urdu.

3. Methodology

The best medium to study the phonology of any language is either its dictionary or native speakers of the target language. As my paper aimed to study all consonant clusters available in Urdu, so dictionary seemed the best medium to study them. For this purpose, I have selected *"Urdu Feroz ul lughaat"* containing more than one lac words, proverbs and idiomatic sentences.

All the root words with "CC" cluster were copied to examine the behavior of the clusters. All the words were transcribed and target templates were identified. Words belonging to English origin were not considered. "[n] has two orthographic representations for nasal. One of which is called *"noon ghunna"* like [hãs] (smile) and other is *"noon"* like [həns] (swine)". In current study, only *noon* sound was considered as *noonghunna* only nasalizes the preceding vowel.

In the first step all words with consonant clusters were searched and enlisted. In the second step, frequency of all clusters was found (see Appendix A). Sounds on basis of manners were also grouped.

4. Results

In *Fero-ul-lughaat*, 699 words having 702 consonant clusters have been found (see Appendix A). None of these was found following the MOP and 242 out 702 clusters were observed violating SSP.
Total templates with CC clusters = 702
The number of templates violating SSP = 242
Other cluster combinations are: Fricative-stops =168, Tap-stops = 077, Both stops = 046, Tap-fricatives = 043, Fricative-taps = 040, Both fricatives and Fricatives-nasals = 039, Stop-fricatives = 038, Stop-taps = 037, Fricatives-liquids = 027, Liquid-stops = 022, Stop-liquids and Stop-nasals = 016, Nasal-fricatives = 015, Tap-nasals = 014, Nasal-stops = 012, Liquid-fricatives and Tap-affricatives = 010, Affricative-stops and Affricative-fricatives = 005, Both nasals = 004, Nasal-liquids, Affricatives-taps, Nasal-taps, Liquids-nasals and Affricative-nasals = 003, Liquids-nasals and Fricative-affricatives = 002, Affricative-liquids and Fricative-retroflex = 001, Both approximants, Both Affricatives, Both Liquids, Both Taps, Nasal affricative, Liquid affricatives, Stop affricatives, Tap liquids and Liquid taps = 000.

As mentioned above these results were bases upon Urdu dictionary *Feroz ul lughaat.*

5. Discussion

Urdu is a language which has the ability to absorb new words in it easily. As Persian, Arabic, Hindi, Sanskrit, Portuguese and other local languages contribute a lot in the vocabulary of Urdu so these languages continue to cast their influence in the direction of evolution of Urdu. The syllable templates of Urdu are similar to the templates of these languages. Each language has its own paradigms and parameters and puts restrictions on its templates and syllabification.

The data observed reveal that Urdu does not follow Maximal Onset Principle (MOP) as it is very sensitive to its onsets and allows only one consonant in this position. Urdu has mixed behaviour towards Sonority Sequence Principle (SSP). It neither follows it completely nor goes against it completely. However, it follows Maximal Coda Principle (MCP) completely.

MCP prefers maximum consonants in the coda position leaving only one in the onset position. There are rare clusters with three consonants in Urdu. Mostly there are two consonants clusters. In case of three consonants, the two will go to the coda of the preceding syllable leaving one to form the onset of following syllable. For example a word [gʊzaʃʈni] has a cluster of three consonants [ʃʈn]. According to this principle, the first two [ʃʈ] will take coda position of the first syllable and [n] will occupy the place of coda of the following syllable like [gʊ.zaʃʈ.ni]

The number of consonant clusters found in the above said dictionary was 185. The cluster [sʈ] had highest frequency 39 and [rdʼ], [xʈ] with 30 and 22 respectively. (For detail see Appendix B)

If we talk about the clusters from articulator point of view then some of the sounds are more frequent than the others. Sometime both consonants in the cluster are voiced and other time voiced-unvoiced combination is occurring mostly. Some other interesting facts have also been found. Fricative-stop clusters have been found in the highest number i.e., 130 out of 168.

If we observe these clusters in a little detail, then we find that in case of both stops, Urdu does not prefer both voiced consonants. Unvoiced-unvoiced combinations are more preferred e.g., out of 168 Fricative-stops, 130 are both unvoiced, both voiced, unvoiced-voiced, voiced-unvoiced are 13, 19 and 06 respectively. Similarly 77 Tap-stops with 47 both

voiced and 30 voiced-unvoiced; 46 both stops with 03 both voiced, 11 both unvoiced, 17 unvoiced-voiced and 15 voiced-unvoiced have been observed.

Urdu does not take voiced fricatives as second consonant, if the first one is tap [r]. Tap [r] also prefers unvoiced before it. In case of both fricatives, both unvoiced are preferred. Nasals [m], [n] also prefer unvoiced preceding fricatives. Both unvoiced stop-fricatives are preferred in clusters. Tap [r] likes unvoiced stops in preceding position. Liquid [l] also prefers unvoiced preceding fricatives and unvoiced stops in both positions. In case of Stop-nasal cluster combination, unvoiced stops are more frequent. Nasals also prefer unvoiced fricatives and unvoiced stops in the following position. Liquid [l] mostly takes unvoiced fricative consonants. Tap [t] prefers voiced affricatives in the following position and most of the affricative-stops clusters are voiced. Unvoiced affricatives do not take unvoiced stops. Nasals are always preceded by voiced affricatives. Taps always take voiced affricatives before them. There are only two examples of fricative-affricative combinations. These show that fricative-affricatives are either both voiced or unvoiced fricative-voiced affricative combination.

Only one example of Affricative-liquid found shows that liquids is followed by voiced affricative. Only one fricative retroflex combination was observed. (See Appendix C)

As for as constraints are concerned, primary data also reveal that [t̪], [tʃ], [ɖ], [v], [ɽ], [ʒ] sounds do not occur in the first position in "CC" clusters where as [ɖ], [ɽ] and [j] do not occur in the second position of these clusters. In Urdu, approximant-approximant, affricative-affricative, liquid-liquid, tap-tap, liquid-affricative, nasal-affricative, stop-affricative and tap-liquid combinations are non-existent. Or we can say that Urdu constrains the following cluster combinations.

None of these clusters has been found in the dictionary. Both approximants, Both Affricatives, Both Liquids, Both Taps, Liquid affricatives, Nasal affricatives, Stop affricatives, Tap liquids, Liquid taps

6. Conclusion

After studying the foresaid dictionary, it was observed that overall 185 consonant clusters within 702 templates were found. The detail study reveals that in Urdu that no consonant cluster is allowed in the onset position. Maximum two consonants are allowed in the coda position (There is one exception [ʔəmrv]

pronounced as [ʔəmr] where [v] is silent). For syllabification, in Urdu only MCP is followed. Mostly there are clusters of unvoiced sounds. Cluster combination of unvoiced fricative-stops is more frequent. If both are stops, voiced stops are least preferred. If there is fricative-tap cluster, then ratio of unvoiced fricatives is higher than voiced ones. Taps do not take voiced fricatives after them.

Unvoiced affricative-stop, affricative-fricative, unvoiced affricative-voiced fricative combinations are not allowed in Urdu. Nasals and taps are always preceded by voiced affricatives. Urdu constrains Approximant-approximant, Affricative-affricative, Liquid-liquid, Tap-tap, Liquid-affricative, Nasal-affricative, Stop-affricative, Tap-liquid, Liquid-tap the following cluster combinations.

References

[1] *Feroz-ul-lughat Urdu*, Feroz Sons Private Limited, 2005.

[2] S. Hussain, *Phonetic Correlates of lexical stress in Urdu*", Northwestern University, IL, USA, 1997.

[3] Ladefoged, *A Course in Phonetics* , Thomson Wadsworth, USA, 2000.

[4] M. Ghazali, "Urdu Syllable Templates", *Annual Report of Center for Research in Urdu Language Processing (CRULP)*, National University of Computer and Emerging Sciences, Lahore, Pakistan, 2002.
Available: www.crulp.org/research/reports/streports02.htm

[5] M. Nazar, "Syllable Templates in Urdu Language", *Annual Report of Center for Research in Urdu Language Processing (CRULP)*, National University of Computer and Emerging Sciences, Lahore, Pakistan, 2002. Available: www.crulp.org/research/reports/streports02.htm

[6] B. Akram, "Analysis of Urdu Syllabification Using Maximal Onset Principle and Sonority Sequence", *Center for Research in Urdu Language Processing (CRULP)*, National University of Computer and Emerging Sciences, Lahore, Pakistan, 2002. Available: www.crulp.org/research/reports/streports02.htm

[7] N. Wyne, "Languages and their families". *Annual Report of Center for Research in Urdu Language Processing (CRULP)*, National University of Computer and Emerging Sciences, Lahore, Pakistan. Available: www.crulp.org/research/reports/streports02.htm

[8] S. Bartlett, G. Kondrak, and C. Cherry, *On the Syllabification of Phonemes* 2009. Available: www.aclweb.org/anthology-new/N/N09/N09-1035.pdf

[9] Hayes,. *Syllables,* 2009.
Available:
www.udel.edu/~heinz/classes/2011/607/materials/…/Hayes2009-13.pdf

[10] A. Saleem, H. Kabir, K. Riaz, M. Rafique, N. Khalid and S. Shahid, "Urdu consonantal and vocalic sounds" *Annual Report of Center for Research in Urdu Language Processing (CRULP)*, National University of Computer and Emerging Sciences, Lahore, Pakistan, 2002. (www.crulp.org/research/reports/streports02.htm)

[11] M. Kenstowicz, *Phonology in Generative Grammar*, Blackwell Publication, UK, 1994.

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

[13] A. McMahon, *An Introduction to English Phonology*, Edinburgh University Press, Edinburgh, 2002.

[14] J. Goldsmith, *Autosegmental and Metrical Phonology*, Basil Blackwell Ltd, UK, 1990.

[15] S. Hussain, "Letter to sound Conversion for Urdu text to speech system", *Annual Report of Center for Research in Urdu Language Processing (CRULP),* National University of Computer and Emerging Sciences, Lahore, Pakistan, 2004.

[16] C. Masiaa, *The Indo-Aryan languages*, Cambridge University Press, Great Britian, 1991.

[17] S. Bokhari, *Phonology of Urdu Language*, Royal Book Company,Karachi, Pakistan, 1985.

[18] *Moraic versus constituent syllables*, Available: http://rnoske.home.xs4all.nl/CV/publicaties/proefschrift/diss2.pdf

[19] *Moraic Phonology*, Available: http://www.ling.fju.edu.tw/phono/farrah/Moraic%20Phonology.htm

Appendix

The appendices are available at the website of the conference. i.e. http://www.cle.org.pk/clt12/