

## Re-syllabification and Multiple Pronunciations in Urdu

Mahwish Farooq<sup>1</sup>, Dr. Asim Mahmood<sup>2</sup>, Dr. Arshad Ali Khan<sup>3</sup>

### ABSTRACT

This is a sequel work deals with the confirmation of phonological rules for re-syllabification and multiple pronunciations of Urdu vocabulary, at larger scale, in speech of Pakistani speakers. The motivation of study is the identification of re-syllabification and multiple pronunciations in 10 hours audio-corpus of a native Pakistani Urdu speaker. This speech corpus has identified 103902 words with different annotations. The observations have identified the multiple pronunciations of same tokens (Farooq & Mumtaz, 2016), (Farooq & Mahmood, 2020). These tokens have same parts-of-speech (POS), spellings and meanings but different pronunciations which ultimately becomes the cause of re-syllabification in different contexts. CLE annotated speech corpus has been used as baseline of this research (Mumtaz, et al., 2014), (Habib, Hijab, Hussain, & Adeeba, 2014). Later, the multiple pronunciations and re-syllabification phenomenon have been cross checked in the speech of 29 native Pakistani Urdu speakers for the identification and confirmation. Consequently, the results of data analysis have also confirmed; the re-syllabification is more context dependent rather than speaker dependent. Moreover, stress is major supra-segmental feature which becomes the reason of multiple pronunciations by causing segmental substitution and attain the status of alternative pronunciations (Farooq & Mumtaz, 2016). Multilingual effect also causes re-syllabification in Urdu lexical items in mental lexicon of speakers. The alternative pronunciations have been triggered by restructuring and re-syllabifications of Urdu vocabulary in speech of 30 Pakistani Urdu speakers. But this paper currently deals with the re-syllabification of Urdu vocabulary depending only on the phoneme or segment alternation. Finally, it is concluded that stress alternation is one major cause to raise re-syllabification and multiple pronunciations on Urdu speech in Pakistan. The multiple pronunciations will identified the rules of ‘word-level phonology’ in Urdu; based on some special rules as exceptions.

**Keywords:** tokens, restructuring, re-syllabification, alternative or multiple pronunciations, phonemic alternation.

---

*University of Management and Technology, Lahore*<sup>1,2</sup>  
*Government College University Faisalabad*<sup>3</sup>

## 1. Introduction

Urdu language belongs to the Indo-Aryan group of languages with more than 100 million speakers around the world. Therefore, Urdu has multiple pronunciations and accents based on geographical regions. Urdu is the National and official language but most importantly a ‘lingua franca’ in Pakistan therefore is esteemed more than any other native language (Farooq, 2015). Currently, the word re-syllabification has been reported and analyzed in Urdu speech for finding out the reasons of multiple pronunciations of Urdu vocabulary. These words or tokens have same parts of speech, spellings and meanings but with different transcriptions. For example, a word احترام (respect /eɦt̪era:m/) (Urdu Lughat: Tarixi Usul Per, 2013) has four more pronunciations i.e., /eɦt̪̤ara:m/, /iɦt̪̤ara:m/, /æɦt̪̤era:m/ and /æɦt̪̤ara:m/ but are interestingly and equally comprehensible by all native Urdu speakers in Pakistan. But these multiple pronunciations may cause difficulty for other language users. These alternative pronunciations (or transcription) may increase the error rate of Natural Language Processing (NLP) tools. Therefore, the motivation of current study is the investigation of phonological reasons behind the re-syllabifications and their multiple pronunciations. Restructuring and multiple pronunciation may appear due to segment alternation, deletion, and addition (Adda-Decker, Boula de Mareüil, Adda, & Lamel) but the current study will only deal with the segment alternation.

A person has an intrinsic connection of sound and meaning which is based on unconscious knowledge of rules. This innate knowledge is used to interpret and produce unknown utterances but there is no ‘habit system or grammar’ which is the main cause for ‘possible phonetic transcriptions’. These variant transcriptions give loophole to the ‘possible phonetic representations’ based on the phonetic features and possible phonotactic conditions in a language (Chomsky & Halle, 1968). Phonological rules are the information of all possible combinations of phonemes in a given language which deal with the alternative pronunciations of a word (Odden, 2005). So, the phonological rules concern with the words’ morphology and morpheme combinations to form meaningful words (Jehsen, 2004).

Different researches have reported the inevitable occurrences of phonological variations in the speech of native speakers (John, NA) which ultimately become the reason of restructuring and re-syllabification of vocabulary. But the re-syllabification may cause confusion among non-

native speakers and also cause errors in machine language. Therefore, 10 hours speech corpus (consisted of 103902 words) of one professional female speaker has been used as a reference point for initiating this research. Afterwards, this wordlist is reused for recording and collecting speech data from 29 more native Urdu speakers. Consequently, this study is done to find out possible reasons for re-syllabification in Urdu vocabulary. Thus, phoneme alternation (Farooq & Mumtaz, 2016), (Hussain, 2005) has been used to explain this research. These phonological variations cause re-syllabification for the alternative pronunciation(s) of the surface forms of already existed phonetic-scripts of Urdu vocabulary. Therefore, a backend strategy is investigated in order to resolve these issues of multiple pronunciations for native and most importantly for non-native speakers. So, the context dependent variations have only catered in this research. Consequently, the alternative pronunciations become the multiple pronunciations of a language.

This paper has proposed Urdu phonological rules for restructuring and re-syllabification of Urdu vocabulary in connected speech. The remaining paper is arranged accordingly; (i) a report on the literature review of phonological rules of restructuring and re-syllabification rules of different languages are discussed in the second section, (ii) third section is about the experimental methodology, (iii) fourth section is about data analysis and results, (iv) fifth section concludes the proposed phonological rules for re-syllabification in Urdu speech, (v) future discussion is presented in section 6 while (vi) section 7 acknowledges the contributions of research participants.

## **2. Literature review**

This research deals with the phonological rules for multiple pronunciations and re-syllabification of Urdu vocabulary in speech of Urdu speakers in Pakistan. A child can unconsciously acquire a language with unlimited ‘phonetic representations’ due to the recursive property of a grammar. This recursive behavior of language will produce number of syntactic and semantic combinations; sometimes even beyond the ‘habit system’ but mostly based on the ‘generalization’ effect (Chomsky & Halle, 1968). When phonemes appear in different contexts, especially in complex syllable structure or even in a sentence, then their articulatory as well as

acoustic attributes may change but in casual style, especially observed in spontaneous conversation (Stevens, 2000).

The phenomenon of re-syllabification has been addressed in different languages by number of researchers. For example, an article has investigated the possibility of re-syllabification for knowing the constraints of Sonority Hierarchy for the segments of [t], and [d]. This case of formal theory has been considered an astray for the empirical researchers (Labov, 1995). Another study has justified lexical and post lexical re-syllabification in the Standard Arabic on Optimality Theory. In the Standard Arabic, restructuring appeared due to the vowel alternation, vowel insertion, vowel deletion, prosthesis effect, and affixation for avoiding inadmissible consonant clusters, especially tri-moraic syllables at lexical and post lexical positions. Moreover, short vowel deletion causes complex coda(s) at word final position by ignoring the Sonority Sequencing Principle (Btoosh, 2019). Another research on 95 Mandarin speakers has also confirmed novel phonological segments and lexical tones in one minute's speech. Their speech has been assessed with different fluency and scientific measures to find the inter-connectedness and cohesion among lexical entities. Such type of multivariate regression analysis has confirmed their multilingual proficiency, status, and fluency. The multilingual relationship between English and Chinese language is the main cause for the restructuring of the phonological mental lexicon of native Chinese speakers. Moreover, the rate of restructuring is directly influenced the random use of English and Chinese languages or dialects (Neergaard, Luo, & Huang, 2019).

Number of studies have confirmed the alternative pronunciations of words in a language due to the native language effect after considering phonological rules (Finch, 2000) e.g., English, Russian, Japanese, Czech, Hungarian, Setswana, Dutch, Finish, and Shona (Panevov & Hana, 2010). According to American National Standards Institute (ANSI), the voice quality depends on the habitual variations of the vocal apparatus of a speaker which causes multiple pronunciations and accent variations based on momentary actions of speech segments (Kreiman, Jody; Sidtis, Diana Vanlancker; Gerratt, Bruce, 2014). Each language has indifferent phonemic inventory (Jehsen, 2004) but may lose their phonemic features in connected speech (Roach, 2009) due to the complex phonological rules (Hall, 2005). Moreover, Vander has highlighted the importance

of; (i) speakers' acoustic behavior and (ii) language change variations due to phonological rules (Hulst H. V., 1979). According to Sound Change Theory (SCT), multiple pronunciations and re-syllabifications are inevitable speech features in connected speech (Ohala, 1980) due to inherent variations of "non-programmed features" to articulate an alternative pronunciation (John, NA), (Odden, 2005). But these phonemic features are not sufficient (Hall, 2005) because auditory transcription cannot duplicate human speech with traditional phonetic symbols for identifying multiple pronunciations of a word. Therefore, the multiple pronunciations and restructuring have been catered in "phonetic grammar" of a language (Odden, 2005) after considering its phonological rules. Though, there are number of phonological rules which cause restructuring and re-syllabifications (Finch, 2000) but this research will only analyze phonemic substitution.

### **2.1. Multiple pronunciations and re-syllabification caused with phonemic substitution**

Phoneme alternation is a primary principle to produce multiple pronunciations and re-syllabification. In a connected speech articulation, the substitution of one speech sound with another is called phonemic substitution (Jehsen, 2004). Different languages have reported number of phonological rules, causing of phonemic substitution and re-syllabification. For example, in Hindi, if a nasal stop comes after an oral vowel it will convert its preceding oral vowel in to a nasal vowel (Trigo, NA). Turkish language has reported syllable final devoicing of a voiced stop (Hulst & Weijer, NA) while Czech and German languages have reported word final devoicing. Some Spanish dialects have reported that a voiced obstruent sound becomes fricatives if comes between two vowels (Panevov & Hana, 2010). In connected Persian speech, /r/ stop may appear as three alternative allophones i.e., [r], [ɾ] and [r̥] depending on the phonological environment (Jehsen, 2004). Lithuanian language has reported voice assimilation as a common phonological rule to produce multiple pronunciations (Odden, 2005).

### **2.2. Urdu phonological rules for re-syllabification and multiple pronunciations**

Number of researches has reported different phonological rules in Urdu connected speech but only at segmental level e.g., (i) assimilation of bilabial, velar and nasal consonants in different contexts, (ii) deletion of glottal fricative /h/ (Hussain, 2005), (iii) deletion of glottal stop /ʔ/ (Nawaz, N.A.), and (iv) epenthesis of /ə/ in consonant clusters (Akram, 2002). However, a recent research has claimed that segmental features are not enough to explain multiple pronunciations.

Therefore, it has reported different phonological factors to cater multiple pronunciations in connected Urdu speech i.e. glottalization, stress alternation, individual segmental features, contextual effect of phonemes in connected speech and restructuring of syllables (Farooq & Mumtaz, 2016) and has proved the initial step for the current study. Therefore, the phenomenon of re-syllabification has been investigated in Urdu speech of 29 more Urdu speakers in order to confirm multiple pronunciations of Urdu vocabulary. Methodology and data analysis have been discussed in the subsequent sections.

### **3. Methodology**

Currently, Urdu phonological rules have been reported to cater re-syllabification and multiple pronunciations in the connected speech of 30 native Urdu speakers in Pakistan. The objective of this study is to identify phonological reasons for words' re-syllabification which may cause multiple pronunciations. Therefore, 10 hours recordings of Urdu speech of one female speaker has been used as a reference to confirm the multiple pronunciations. She is a professional vocalist at the Pakistan Radio Station, Lahore and her age is between 35-40 years old. Later, the phenomenon of re-syllabification and multiple pronunciation have been identified and confirmed in Urdu speech of 29 male-female Urdu speakers. They have claimed Urdu as their first language and good comprehension of English language. They have been selected conveniently from different public sector universities of Pakistan. Their age is between the range of 18-25 years and they have completed their higher secondary school education and will graduates in different disciplines. Their speech data has been recorded at 8 KHz in an echoic chamber within PRAAT software. Speech segmentation and annotation has been done at different tiers by using Case Insensitive Speech Assessment Phonetic Alphabets (CISAMPA) in PRAAT (Mumtaz, et al., 2014). The results of data analysis are reported in section 4 for removing confusions. The results have confirmed different reasons for the multiple pronunciations; multilingual effect, speaker dependent variations and contextual variations.

### **4. Results**

Multiple pronunciations and re-syllabification of Urdu vocabulary is the main objective of this research. The ten hours speech corpus has confirmed the transcription duplications at 13717 times with different reasons (Farooq & Mumtaz, 2016). But only 7112 duplicates have been used

as a text corpus and ultimately for speech corpus of this research after considering the scope of the research. This data analysis has been used for the assurance of the re-syllabification for alternative pronunciations of Urdu vocabulary at larger scale. These duplicates have given information about the multiple instances of Urdu vocabulary with similar spellings. This list also contains the information about; (i) transcriptions, (ii) POS tags, (iii) syllables' count of vocabulary, (iv) stress patterns and (v) file IDs. There are different reasons for the mismatches of Urdu word duplicates. It may have; (i) errors in annotation, (ii) homographs, (iii) homophones, (iv) variation in the stress patterns of duplicates, and (v) alternative pronunciations as well as transcription. According to the research demand; first three types are ignored but only fourth and fifth categories are considered for the confirmation of alternative pronunciation. The standard transcription is consulted with "Urdu Lughat: Tarixi Usul Per" (Urdu Lughat: Tarixi Usul Per, 2013) and English meanings of duplicates are incorporated with the consultation of Oxford Urdu-English Dictionary (Parekh, 2013). The data analysis and results of duplicates are reported in the table 1;

**Table 1**

| Phonemic Alteration                         |     |     |                                                         |     |                                                  |                                              |     |
|---------------------------------------------|-----|-----|---------------------------------------------------------|-----|--------------------------------------------------|----------------------------------------------|-----|
| Mutual Alteration of Short Vowels<br>T=5060 |     |     | Mutual Alteration of<br>Short and Medial Vowel<br>T=916 |     | Mutual Alteration of Medial<br>Vowels<br>T = 888 | Mutual Alteration of Long<br>Vowels<br>T=248 |     |
|                                             | ا→آ | و→ا | ا→ا                                                     | ا→ا | ا→ا                                              | ا↔ا                                          | ا↔ا |
| <b>TW</b>                                   | 850 | 850 | 830                                                     | 165 | 293                                              | 440                                          | 124 |
| <b>ST</b>                                   | 550 | 300 | 779                                                     | 95  | 154                                              | 124                                          | 28  |
| <b>AP</b>                                   | 300 | 550 | 51                                                      | 70  | 139                                              | 324                                          | 96  |

Note: **TW** = Total words, **ST** = standard Transcription, **AP** = alternative pronunciations

Later, for the confirmation of suggested variations, 29 male-female Urdu speakers are selected conveniently from public sector universities of Pakistan. They have at least completed their higher secondary school education. Results have confirmed the presence of alternative pronunciations in Urdu speech of 29 Urdu speakers of public sector universities in Pakistan. The current work is based on a hypothesis that Urdu phonological rules directly influence re-syllabification by causing restructuring therefore known as major cause of multiple pronunciations in Urdu speech of Pakistani Urdu speakers. This wordlist is recorded after embedding in different sentences for avoiding stress, boundary effect and confirming contextual

effect. Subsequently, the results have confirmed alternative pronunciations of the given wordlist. This collected data also has same spellings and, parts-of-speech, but with different stress patterns and transcriptions which become the reasons of re-syllables and multiple pronunciations. These multiple instances have been entered in an excel log-sheet for reconfirmation in order to avoid inconsistency and human errors in annotation. Log-sheet also saves the time by making record of each alternative pronunciation with specific file ID. Then, multiple pronunciations have been used only after consulting annotation errors. Alternative pronunciations may occur due to phonemic substitution, stress alternations, epenthesis and ellipsis by causing re-syllabifications of Urdu unique tokens. These multiple pronunciations give broader perspective to reach a conclusive decision about re-syllabification of Urdu vocabulary. This study will ultimately relax the concept of mispronunciations by accepting the alternative pronunciations as alternative variety. Results of data analysis are given in table 2 for making clarity about the context dependent or speaker dependent variations.

Table 2

|      | Total Number of Alternative Words = 3339     |       |       |                                                            |       |                                                 |                                               |
|------|----------------------------------------------|-------|-------|------------------------------------------------------------|-------|-------------------------------------------------|-----------------------------------------------|
|      | Vowel Alternation                            |       |       |                                                            |       |                                                 |                                               |
|      | Mutual Alternation of Short Vowels<br>T=1384 |       |       | Mutual Alternation of Short<br>and Medial Vowels<br>T=1012 |       | Mutual Alternation of<br>Medial Vowels<br>T=478 | Mutual Alternation<br>of Long Vowels<br>T=465 |
|      | ə → ɪ                                        | o → ə | ɪ → ə | e → ə                                                      | e → ɪ | e → æ                                           | e: → æ:                                       |
| SP 1 | 13                                           | 15    | 10    | 7                                                          | 13    | 24                                              | 06                                            |
| SP2  | 13                                           | 15    | 11    | 17                                                         | 12    | 20                                              | 12                                            |
| SP3  | 24                                           | 14    | 11    | 11                                                         | 19    | 13                                              | 22                                            |
| SP4  | 14                                           | 18    | 12    | 23                                                         | 15    | 13                                              | 11                                            |
| SP5  | 17                                           | 22    | 10    | 22                                                         | 11    | 16                                              | 21                                            |
| SP6  | 18                                           | 20    | 11    | 12                                                         | 16    | 18                                              | 18                                            |
| SP7  | 19                                           | 22    | 13    | 11                                                         | 17    | 10                                              | 13                                            |
| SP8  | 22                                           | 20    | 11    | 12                                                         | 15    | 10                                              | 12                                            |
| SP9  | 5                                            | 11    | 19    | 14                                                         | 14    | 10                                              | 14                                            |
| SP10 | 11                                           | 12    | 15    | 10                                                         | 15    | 13                                              | 11                                            |
| SP11 | 19                                           | 14    | 14    | 10                                                         | 11    | 16                                              | 19                                            |
| SP12 | 11                                           | 19    | 14    | 14                                                         | 10    | 14                                              | 20                                            |
| SP13 | 12                                           | 15    | 10    | 15                                                         | 13    | 11                                              | 22                                            |
| SP14 | 11                                           | 12    | 15    | 10                                                         | 15    | 13                                              | 11                                            |
| SP15 | 19                                           | 14    | 14    | 10                                                         | 11    | 16                                              | 19                                            |
| SP16 | 14                                           | 18    | 12    | 23                                                         | 15    | 13                                              | 11                                            |
| SP17 | 17                                           | 22    | 10    | 22                                                         | 11    | 16                                              | 21                                            |
| SP18 | 14                                           | 18    | 12    | 23                                                         | 15    | 13                                              | 11                                            |
| SP19 | 17                                           | 22    | 10    | 22                                                         | 11    | 16                                              | 21                                            |

|           |     |     |     |     |     |     |     |
|-----------|-----|-----|-----|-----|-----|-----|-----|
| SP20      | 15  | 16  | 17  | 17  | 25  | 15  | 14  |
| SP21      | 13  | 7   | 7   | 19  | 18  | 25  | 19  |
| SP22      | 15  | 16  | 17  | 17  | 25  | 15  | 14  |
| SP23      | 13  | 7   | 7   | 19  | 18  | 25  | 17  |
| SP24      | 19  | 17  | 17  | 23  | 22  | 22  | 14  |
| SP25      | 18  | 11  | 11  | 20  | 22  | 12  | 15  |
| SP26      | 7   | 23  | 23  | 23  | 23  | 24  | 11  |
| SP27      | 17  | 21  | 22  | 22  | 22  | 14  | 23  |
| SP28      | 11  | 24  | 24  | 19  | 24  | 18  | 11  |
| SP29      | 23  | 13  | 23  | 19  | 23  | 19  | 21  |
| SP30      | 22  | 18  | 23  | 20  | 25  | 14  | 11  |
| TP        | 463 | 496 | 425 | 506 | 506 | 478 | 465 |
| Std. Dev. | 4.5 | 4.5 | 4.8 | 5.1 | 4.9 | 4.4 | 4.6 |

## 5. Data analysis and discussion

Urdu phonological rules have suggested re-syllabification with sound change rules. Re-syllabification is primary reason for alternative/multiple pronunciations of already existed phonetic scripts of different surface forms. The results have confirmed three main causes for re-syllabification and restructuring of Urdu vocabulary. Those are phonemic alternation, ellipsis, and epenthesis. But in the subsequent sections, this research will only discuss phonemic alternation especially vowel substitution in Urdu vocabulary.

### 5.1.1. Re-syllabification and Phonemic alternation

Mutual shuffling of speech sounds with other phonemes is called phonemic alternation and these variant forms are called ‘alternants’. In general, an alternation rule can be written as, A converts into B in X to Y context;  $A \rightarrow B/X\_Y$ . It is considered the primary principle for re-syllabification and multiple pronunciations. Though, morphology does not allow ‘alternants’ but phonology accepts them with the contextual information of a segment which can alternate stereotypical features of speech segments;

“....every morpheme has only one phonological form. Any variation in the phonetic shape of a morpheme results from the operation of regular phonological rules”.

(Jehsen, 2004).

Urdu phonology has also identified various ‘alternants’. But a native speaker can articulate one ‘alternant’ at a single instance of time. Currently, according to the data analysis, Pakistani

Urdu speakers can cause re-syllabification by articulating multiple pronunciations of vocabulary with the substitution of one vowel with another vowel. There are four different types of vowel substitutions; (i) mutual alternation of short vowels, (ii) mutual alternation of short and medial vowels, (iii) mutual alternation of medial vowels, and (iv) mutual alternation of long vowels. The reasons of first two types of alternations are avoided to discuss because they are speaker dependent variations therefore are out of the scope of this research.

### 5.1.1. Re-syllabification and mutual alternation of short vowels

First condition is the mutual alternation of short vowels; it may occur when one short vowel alternates with another short vowel e.g. in the word  $\text{ﮬﻤﻢ}$  (proper name /mohəmməd/), /ʊ/ is substituted either with a short vowel /ə/ or a medial vowel /o/ to form two more alternative pronunciations i.e., /mohəmməd/ and /mohəmməd/. In another example, the word  $\text{ﺑﻠﻨﺪ}$  (high /bələnd/) may articulated with two alternative pronunciations i.e., /bələnd/ and /bələnd/ by substituting /ə/ with short vowel /ʊ/ or medial vowel /o/ respectively. According to the data analysis, phoneme alternation is occurred at two different levels i.e. at word initial and word medial positions. These phonemic alteration follows some specific pattern; front vowels substituted with the front vowel and back vowel alternated with the back vowels but possibly nearest vowel on quadrilateral i.e., /ɪ/ is substituted either with /e/, /æ/, or /ə/. Secondly, it is also observed that shuwa /ə/ is the only short vowel which is randomly appeared regardless of the position of substituted vowel on quadrilateral. Its central position on the quadrilateral vowel chart can easily describe the main cause of random substitution of vowels with /ə/ short vowel. The reason is; tongue has to cover a minimum distance for its articulation therefore, occurs randomly substituted with other oral vowel or even causes epenthesis in complex consonant clusters in the speech files of Urdu native speakers in Pakistan. Thirdly, a word medial schwa /ə/ is substituted with /æ/ medial vowel in the speech reference files. The reason of first and second types of segment alternations have proved speaker dependent variations because haven't followed specific contextual pattern. On the other hand, third type of substitution is contextual variations. In Urdu speech of Pakistani Urdu speakers, some common conditions have been observed for the phonemic alternations. Those are;

1. Phoneme alternation has been observed only in content words either monosyllabic, disyllabic, or multisyllabic.
2. A mutual alternation of short vowels occur if the word starts with glottalization effect or syllabic stress. In general, we can write the rule as;  $A \rightarrow B / \_ \# [ + \text{glottalization} ]$  while more specifically as;

**(i)  $e \rightarrow \text{æ} / \_ \# [ + \text{glottalization} ]$**

$$\begin{pmatrix} +\text{highmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} / - \begin{pmatrix} +\text{syll.} \\ \# - \text{back} \end{pmatrix} -\text{back} + \text{C.G.}$$

**(ii)  $\text{u} \rightarrow \text{o} / \_ \# [ + \text{glottalization} ]$**

$$\begin{pmatrix} +\text{highmid} \\ +\text{short} \\ +\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ +\text{round} \end{pmatrix} / - \begin{pmatrix} +\text{syll.} \\ \# + \text{back} \end{pmatrix} + \text{back} + \text{C.G.}$$

3. A stressed articulation mostly become the cause to alternate a short vowel with another short vowel but without changing the place of articulation. Means that, a front vowel will only alternate with front and back vowels alternate with back vowels. But only shuwa vowel plays an important role by appearing in place of both short vowels ( $\text{u}$  and  $\text{i}$ ) due to its central position in quadrilateral. Generally, the rule can be written as;  $A \rightarrow B / \_ \# [ + \text{stress} ]$  while more specifically as;

**(iii)  $e \rightarrow \text{æ} / \_ \# [ + \text{stress} ]$**

$$\begin{pmatrix} +\text{highmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} / - \begin{pmatrix} +\text{syll.} \\ \# - \text{back} \end{pmatrix} -\text{back} + \text{stress}$$

**(iv)  $\text{u} \rightarrow \text{o} / \_ \# [ + \text{stress} ]$**

$$\begin{pmatrix} +\text{highmid} \\ +\text{medial} \\ +\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ +\text{round} \end{pmatrix} / - \begin{pmatrix} +\text{syll.} \\ \# + \text{back} \end{pmatrix} + \text{back} + \text{stress}$$

4. If the word starts with glottalization or syllabic stress, the long vowel alternates with another long or short vowel. In general, we can write the rule as;  $A \rightarrow B$  /\_#[+glottalization] while more specifically as;

**(v)  $e: \rightarrow \text{æ}:$  /\_#[+glottalization]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{long} \\ -\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \#-\text{back} \end{pmatrix} \text{ -back + C.G.}$$

**(vi)  $\text{ɔ}:\rightarrow \text{o}:$  /\_#[+glottalization]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ +\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{long} \\ +\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \#+\text{back} \end{pmatrix} \text{ + back + C.G.}$$

**(vii)  $\text{ɔ}:\text{o}:\rightarrow \text{ʊ}$  /\_#[+glottalization]**

$$\begin{pmatrix} +\text{mid} \\ +\text{long} \\ +\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{highmid} \\ -\text{long} \\ +\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \#+\text{back} \end{pmatrix} \text{ + back + C.G.}$$

**(viii)  $e: \rightarrow \text{æ}$  /\_#[+glottalization]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \#-\text{back} \end{pmatrix} \text{ -back + C.G.}$$

5. Stressed articulation also becomes the cause for the mutual alternation of long vowels but without changing their place of articulation. Generally, the rule can be written as;  $A \rightarrow B$  /\_#[+stress] while more specifically as;

**(ix)  $e: \rightarrow \text{æ}:$  /\_#[+stress]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{long} \\ -\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \#-\text{back} \end{pmatrix} \text{ -back + stress}$$

**(x)  $\text{ɔ}:\rightarrow \text{o}:$  /\_#[+glottalization]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{long} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \end{pmatrix}$$

+round                      +round                      #+back    + back + stress

**(xi) ɔ:/o: → ʊ / \_#[+glottalization]**

$$\begin{pmatrix} +\text{mid} \\ +\text{long} \\ +\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{highmid} \\ -\text{long} \\ +\text{round} \end{pmatrix} \not\rightarrow \begin{pmatrix} +\text{syll.} \\ \#+\text{back} \end{pmatrix} + \text{back} + \text{stress}$$

**(xii) e: → æ / \_#[+stress]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ -\text{long} \\ -\text{round} \end{pmatrix} \not\rightarrow \begin{pmatrix} +\text{syll.} \\ \#-\text{back} \end{pmatrix} -\text{back} + \text{stress}$$

6. If a word is articulated in stressed utterance then a diphthong may substitute with another phoneme by causing syllable breakage by inserting an approximant e.g., a word کُیا [kea:] is alternatively pronounced as [keja:] i.e., diphthong breakage with /j/ insertion.

**5.1.2. Re-syllabification and mutual alternation of short and medial vowels**

Second condition is the mutual alternation of short and medial vowels. According to the data analysis, such alternations mutually appear when a short vowel alternates with a medial vowel and vice versa e.g. the word ارتکاب (committing an offence) is articulated in three alternative forms i.e., one standard form (/ɪr̥t̪ek̪a:b/) and two alternative forms (i.e., /ɪr̥t̪u̠ka:b/ and /ɪr̥t̪əka:b/) with the substitution of a medial vowel [e] into a short vowel [ə] or [ɪ]. Another Urdu word محمد (proper noun) is articulated with three different pronunciations; (i) one standard form (/mʊhəmməd/) and two alternative but mutually understandable pronunciations (i.e., /mohəmməd/ and /məhəmməd/) with the substitution of a short vowel [ʊ] into a medial [o] or short vowel [ə] respectively. In first word, a medial vowel is mutually alternated with a short vowel [ɪ/ə] and in second word, a short vowel /ʊ/ alternates with a medial vowel [o] which become the cause of re-syllabifications.

**5.1.3. Re-syllabification and Mutual alternation of medial vowels**

Thirdly, the mutual alternation of medial vowels also causes re-syllabification of Urdu vocabulary. This condition comes true especially in polysyllabic Urdu vocabulary which starts with ح، الف letters e.g., the word، احتیاط (care) is articulated with two different pronunciations i.e., one standard form /eht̪t̪ja:t̪/ and one alternative form /æht̪t̪ja:t̪/. Stressed articulation is the second

and glottalization is the third most important condition for the substitution of /e/ medial vowel into /æ/ medial vowel. If Urdu vocabulary has ح الف, ح letters at word initial position than a speaker can articulate it either at a normal speaking rate or as a stressed articulation. So, if a speaker utters a word with normal speaking rate than s/he pronounces that vocabulary with /e/ medial vowel which is supposed to be a standard form but alternates medial vowel /e/ with /æ/ only when s/he articulated the same vocabulary with stressed condition. Moreover, a stressed or glottalized version of the vocabulary also shows larger speech segment than the normal speech segments. Conditions for the alternative pronunciation can be describe as;

- i. the words; احترام (respect /eħtəra:m/), احتجاج (protest /eħtədʒa:dʒ/), احرام (an unstitched white cloth for Hajj /ehra:m/), احساس (feeling /ehsa:s/), احسان (good deed /ehsa:n/), احكام (orders /ehka:m/), احتما (phlebotomy /eħtəma:m/), have been articulated with alternative pronunciations e.g., /æħtəra:m/, /æħtədʒa:dʒ/, /æhra:m/, /æhsa:s/, /æhsa:n/, /æhka:m/, /æħtəma:m/ respectively. But this alternation appears only when all these words start with glottalization then /e/ medial vowel is articulated with substituted medial vowel /æ/.

$$e \rightarrow \text{æ} / \_ \# [+glottalization]$$

$$\begin{pmatrix} +\text{highmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \# -\text{back} \end{pmatrix} -\text{back} + \text{C.G.}$$

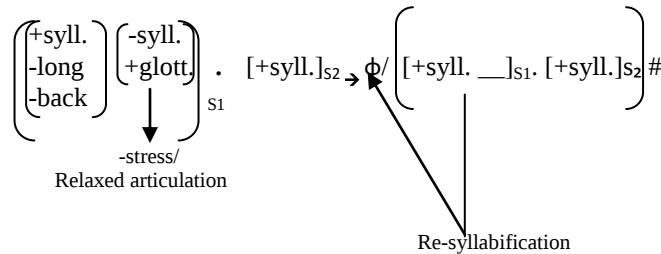
- ii. secondly, the words; احترام, احتجاج, احرام, احساس, احسان, احكام, احتما have been articulated within stressed speech than /e/ medial vowel is alternated with /æ/ medial vowel by giving alternative pronunciations of these words e.g., /æħtəra:m/, /æħtədʒa:dʒ/, /æhra:m/, /æhsa:s/, /æhsa:n/, /æhka:m/, /æħtəma:m/, respectively.

$$e \rightarrow \text{æ} / \_ \# [+stress]$$

$$\begin{pmatrix} +\text{highmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{medial} \\ -\text{round} \end{pmatrix} / \begin{pmatrix} +\text{syll.} \\ \# -\text{back} \end{pmatrix} -\text{back} + \text{stress}$$

#### 5.1.4. Re-syllabification and Mutual Alternation of medial and long vowels

Mutual alternation of medial and long vowels is the fourth situation for re-syllabification and multiple pronunciations of Urdu vocabulary. But the condition is the co-occurrence of الف and ع at word initial and medial positions. Data analysis has confirmed the conversion of a medial vowel into a long vowel and the reason is the relaxed or unconscious articulation of Pakistani Urdu speakers. Because in order to fill the time slot of a syllable, the deletion of glottal stop ع [ʔ] converts its preceding medial vowel into a long vowel (Nawaz, Deletion Rules in Urdu Language, N.A.). For example, at normal speaking rate, the word اعتبار (Trust /eʔt̪eba:r/) is articulated as /e:t̪eba:r/, اعجاز (miracle /eʔd̪ʒa:z/) is articulated as /e:d̪ʒa:z/, اعتماد (Trust /eʔt̪ema:d̪/) is pronounced as /e:t̪ema:d̪/, اعتراض (objection /eʔt̪era:z/) is alternatively pronounced as /e:t̪era:z/, اعلان (to announce /eʔla:n/) is articulated as /e:la:n/. So, phonologically it will be right if we write this re-syllabification rule as;



[.] a dot stands for a syllable boundary and # stands for a word boundary

#### 5.1.5. Re-syllabification and Mutual Alternation of Long vowels

Mutual alternation of long vowels is the fourth situation for the alternative pronunciation of Urdu vocabulary but if the articulated words are found in stress articulation e.g., word تینیس (twenty three) has two pronunciations; one is standard /t̪e:i:s/ while /t̪æ:i:s/ as an alternative pronunciation, word گیند (ball) has two randomly used pronunciations; one is standard /ge:nd̪/ and alternative pronunciation /gæ:nd̪/, and the word چڑیل (witch) has two pronunciations; one standard /t̪ʊɾe:l/ and the other is alternative /t̪ʊɾæ:l/. Data analysis has also confirmed the conversion of a medial vowel into a long vowel after deleting glottal stop ع [ʔ] but the condition is the co-occurrence of الف and ع at word initial position e.g. in the word اعتبار (Trust /eʔt̪eba:r/)

such type of words have standard transcription with /e/ medial vowel but this medial /e/ vowel converted into long vowel /e:/ with the deletion of glottal stop /ʔ/ because Pakistani Urdu speakers are eliminated it at any position (Nawaz, N.A.). This would not be wrong if we take /e:/ long vowel as a standard segment therefore in this study glottal stop /ʔ/ consciously has been deleted in the alternative transcriptions of the words. Presently, we are dealing with the mutual alternation of /e:/ long vowel in /æ:/ long vowel which became the cause of multiple pronunciations in polysyllabic words e.g.,

- i. If the vowel has prevocalic glottalization;

**e: → æ: / \_ [+glottalization]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ -\text{round} \\ -\text{back} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{long} \\ -\text{round} \\ -\text{back} \end{pmatrix} / \_ \begin{pmatrix} +\text{syll.} \\ +\text{C.G} \end{pmatrix}$$

- ii. Due to the stressed articulation; long vowel /e:/ would be substituted with the long vowel /æ:/

**e: → æ: / \_ [stress]**

$$\begin{pmatrix} +\text{highmid} \\ +\text{long} \\ -\text{round} \\ -\text{back} \end{pmatrix} \rightarrow \begin{pmatrix} +\text{lowmid} \\ +\text{long} \\ -\text{round} \\ -\text{back} \end{pmatrix} / \_ \begin{pmatrix} +\text{syll.} \\ +\text{stress} \end{pmatrix}$$

## 6. Conclusion and future discussion

Finally, the speech analysis of Urdu corpus has confirmed the multiple pronunciations occur due to the re-syllabification of Urdu vocabulary. The results have also confirmed Urdu phonological rules at larger scale. All the above mentioned rules have been confirmed after collecting the consent of native Urdu speakers in Pakistan. (i) Re-syllabification is observed only in open class words i.e. nouns, adjectives, verbs, etc. (ii) Stress play important role in re-syllabification of Urdu vocabulary e.g., unstressed articulation causes elision and stressed articulation causes epenthesis. (iii) Phonemic elision has reported only in multisyllabic words.

Presently, dictionaries only incorporate morphological information but phonological information of vocalic alternation is not the part of any dictionary. Therefore, these suggestive variation should be incorporated in new dictionaries in order to cater alternative pronunciations of Urdu vocabulary. These phonological information will prove beneficial for speech recognition and natural language processing tools for avoiding speech errors. It will ultimately improve the quality of artificial languages and NLP tools.

## 7. Acknowledgement

This paper is extracted from my Ph.D. research. I am obliged to my supervisor for his valuable guidance and also thankful to all volunteer speakers who help me out in finding the solution of my hypothesis.

## 8. References

- Adda-Decker, M., Boula de Mareüil, P., Adda, G., & Lamel, L. (n.d.). Investigating Syllabic Structure and its Variations in Speech from French Radio Interviews. *LIMSI-CNRS, BP 133*.
- Akram, B. (2002). Analysis of Urdu Syllabification Using Maximal Onset Principle and Sonority Sequence Principle. (pp. 160-166). Center for Research in Urdu Language Processing.
- Bhatti, R. (2016). Identification of Diphthongs in Urdu and their Properties. *Conference for Language and Technology, CLT16*. Lahore.
- Btoosh, M. A. (2019, December). Resyllabification in Standard Arabic: A Constraint-Based Approach. *SKASE Journal of Theoretical Linguistics*(, 16(2). Retrieved July 27, 2020
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York, The United States of Aamerica: Harper & Row, Publishers. Retrieved July 1, 2020
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. 49 East 33rd Street, New York, N.Y. 10016, The United State of America: Harper & Row Publishers.

- Farooq, M. (2015). *An Acoustic Phonetic Study of Six Accents of Urdu in Pakistan* (Vol. Thesis). University of Management and Technology, Johar Town, Lahore, Pakistan.
- Farooq, M., & Mahmood, A. (2020, June). Acoustic Effect of Urdu Phonological Rules on English Connected Speech (Unpublished). *Journal of Language and Literature Review (LLR)*.
- Farooq, M., & Mumtaz, B. (2016). Urdu Phonological Rules in Connected Speech. *Conference of Language and Technology (CLT16)*. Lahore: Center for Language Engineering, University of Engineering and Technology, Lahore.
- Finch, G. (2000). Phonetics and Phonology. In J. Peck, & M. Coyle (Eds.), *Linguistic Terms and Concepts* (1st ed., Vol. 1, pp. 33-76). New York: St. Martin's Press, INC.
- Habib, W., Hijab, R., Hussain, S., & Adeeba, F. (2014). Design of Speech Corpus for Open Domain Urdu Text to Speech System Using Greedy Algorithm. *Conference on Language and Technology (CLT14)*. Karachi, Pakistan.
- Hall, C. K. (2005). Defining Phonological Rules over Lexical Neighbourhoods: Evidence from Canadian Raising. In J. Alderete, C. -h. Han, & A. Kochetov (Ed.), *Proceeding of the 24th West Coast Conference on Formal Linguistics: Somerville MA: Cascadilla Proceedings Project* (pp. 191-199). USA: Cascadilla Press.
- Hulst, H. V. (1979). Rule Conversion in Phonology. *Dutch Lexicological Institute Leiden*, 336-349.
- Hulst, H. v., & Weijer, J. v. (N.A.). Topics in Turkish Phonology.
- Hulst, H. v., & Weijer, J. v. (NA). Topics in Turkish Phonology.
- Hussain, S. (2005). Phonological Processing for Urdu Text to Speech System. *Localisation in Pakistan, in Localisation Focus: The International Journal for Localisation*, 3(4).

- Jehsen, J. T. (2004). Distinctive Features. In J. T. Jehsen, & K. E. Korener (Ed.), *Principles of Generative Phonology* (iv ed., pp. 79-106). The Netherlands, USA: John Benjamins Publishing Company, Amsterdam/Phila Delphia.
- John, O. J. (NA). The Application of Phonological Universals in Speech Pathology. Phonology Laboratory, Department of Linguistics, University of California, Berklay, California.
- John, O. J. (NA). The Application of Phonological Universals in Speech Pathology. Phonology Laboratory, Department of Linguistics, University of California, Berklay, California.
- Kahn, D. (1976, August 9). Syllable Based Generalizations in English Phonology. *1, 1*, 218. (P. Kiparsky, Ed.) Massachusetts Institute of Technology.
- Kreiman, Jody; Sidtis, Diana Vanlancker; Gerratt, Bruce. (2014, June). Defining and Measuring Voice Quality. *Sound to Sense* (pp. 163-168). New York, USA: MIT.
- Labov, W. (1995, September). *Resyllabification*. Retrieved July 27, 2020, from Proceedings of the International Workshop on Language Variation and Linguistic Theory, Nijmegen: <https://www.ling.upenn.edu/~wlabov/Papers/Resyllab/Resyllabification.html>
- Mendoza, D. R. (N.A.). Phonological Progresses. Philippines. Retrieved September 28, 2015
- Mendoza, D. R. (n.d.). Phonological Progresses. Philippines. Retrieved September 28, 2015
- Mumtaz, B., Hussain, A., Hussain, S., Mahmood, A., Bhatti, R., Farooq, M., & Rauf, S. (2014). Multitier Annotation of Urdu Speech Corpus. *CLT14 - 5th Conference on Language and Technology*. Karachi.
- Murcia, M. C., Brinton, D. M., & Janet, M. (2010). *Teaching Pronunciation Hardback with Audio CDs (2): A Course Book and Reference Guide* (2nd, illustrated, reprint ed.). Cambridge University Press.
- Nawaz, S. (n.d.). *Deletion Rules in Urdu Language*. Center for Research in Urdu Language Processing, CRULP, Lahore.

- Nawaz, S. (N.A.). *Deletion Rules in Urdu Language*. Center for Research in Urdu Language Processing, CRULP, Lahore.
- Neergaard, K. D., Luo, J., & Huang, C.-R. (2019, November 05). Phonological network fluency identifies phonological restructuring through mental search. *Scientific Reports*. doi:<https://doi.org/10.1038/s41598-019-52433-w>
- Nierfeld, G., & Theo, V. (1974). Restructuring. *Lingua*, 33, 137-156.
- Nierfeld, G., & Theo, V. (1974). Restructuring. *Lingua*, 33, 137-156.
- Odden, D. (2005). Feature Theory. In D. Odden, *Introducing Phonology* (pp. 129-168). Cambridge, , United States, NewYork: Cambridge University Press.
- Ohala, J. J. (1980). The Application of Phonological Universals in Speech Pathology. *Speech and Language: Advances in Basic Research and Practice* , 3(0-12-608603-6), 75-94.
- Ohala, J. J. (1993). Coarticulation and Phonology. *Language and Speech*, 2(36), 155-170.
- Panevov, J., & Hana, J. (2010, October 13). Intro to Linguistics \_ Phonology. 1-5.
- Panevova, J., & Hana, J. (2010, October 13). *Intro to Linguistics - Phonology*. Retrieved September 28, 2015, from <https://ufal.mff.cuni.cz/~hana/teaching/2013wi-ling/04-Phonology.pdf>
- Parekh, R. (Ed.). (2013). Oxford Urdu-English Dictionary. 1, 1, 1165. (S. M. Saleem ud Din, & S. Anjum, Compilers) Pakistan: Oxford Publication.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* (4th ed., Vol. 1). Cambridge University Press.
- Stevens, K. N. (2000). *Acoustic Phonetics*. Hong Kong: Massachusetts Institute of Technology. Retrieved July 2, 2020
- Trigo, P. (N.A.). LX 513 Phonology. *Hindi (GP)*, 175-76.
- Trigo, P. (NA). LX 513 Phonology. *Hindi (GP)*, 175-76.

Urdu Lughat: Tarixi Usul Per. (2013, June). *Urdu Lughat: Tarixi Usul Per, 1-22, 3rd.* (D. M. Haq, D. A.-u.-L. Siddiqi, D. S. Sabzvari, M. N. Amrohvi, & M. Arif, Compilers) Muhit e Urdu Press, Urdu Lughat Board, Karachi.

Waqar, A., & Waqar, S. (2002). Identification of Diphthongs in Urdu and their Acoustic Properties. (pp. 16-26). Lahore, Pakistan: Center for Language Engineering, University of Engineering and Technology, Lahore.