**Constraints on Coda Consonant Clusters in Actual Persian Speech in Accordance with Universal Patterns**

**Abstract**

Enjoying support from both acoustic and typological research, the Sonority Sequencing Principle (1984) and the Syllable Contact Law are among the best-known explanations suggested for patterns observed about syllable structures. It is stated that universally, syllable onsets tend to gradually increase in sonority, and, conversely, syllable codas tend to gradually decrease in sonority. Several pieces of research have been performed to verify this in Persian. None has ever based its analyses on actual Persian speech, however, where the syllable structure displays different behavior, as most phonological processes that characterize connected speech are not observed in discrete words. The present paper investigates the limitations Persian consonant clusters in the coda position of syllables are subject to. These syllables are those extracted from actual connected speech rather than taken from isolated entries of a Persian dictionary. It was found that universal patterns were obeyed in about 195 (56.7 percent) of all the 344 syllables with unequally sonorant coda consonants, while the other 149 (43.3 percent) disobeyed the principle. Moreover, 217 (38.6 percent) of all the 561 syllables examined contained homogeneous coda consonant clusters, where rising or falling sonority would not make sense.

**Keywords:** Sonority Sequencing Principle; Syllable Contact Law; Coda; Cluster.

**1. Introduction**

There are universal patterns of the linear order of segments in a syllable. Vowels constitute the most sonorant group of segments, followed respectively by glides, liquids, nasals, and obstruents on the sonority hierarchy. The universal principle of sonority sequencing, abbreviated as SSP, and the rather similar law of syllable contact, abbreviated as SCL, predict that sonority decreases from the syllable core to both its edges (Selkirk, 1984).

As far as Persian is concerned, it is only the CVCC syllable that contains a cluster, since this is the maximal one based on Persian phonotactics. Onsets containing consonant clusters are unattested, and the number of consonants allowed to appear in the rhyme is no more than two. Therefore, the sonority sequencing principle in Persian would only require these syllables to end in a consonant less sonorant than its neighbor. To examine whether or not this is the case, a sufficient amount of data is to be analyzed.

Linguistic universals have been the subject of intense research for decades both in the Chomskyan and the Greenbergian traditions. This has included all branches including phonology. Observation of data from a variety of languages from different linguistic genera and areas has confirmed the sonority sequence profile in many cases. There is not a reason to expect Persian to be an exception. Although this has been considered before, it deserves further attention using more authentic data.

Phonological processes like liaison, moving the final coda consonant of a syllable to function as the initial onset consonant of the next, and elision, where a segment is deleted from a segment sequence, are more likely to occur in more authentic data (and not even expected in static data, like those taken from a lexicon). Clearly, this can greatly affect CVCC syllables, primarily by reducing them to CVC syllables.

For example, the Persian noun *fekr*, meaning ‘idea’, represents a CVCC string in the original dictionary form, hence enumerated as an instance of such syllables in a lexicon-based investigation of an issue concerning the phonological structure. However, the same noun is pronounced as *fek* within the genitive construction *fekr-e xub*, meaning ‘good idea’, where the second coda consonant is pronounced together with the *ezafe* enclitic vowel rather than with the preceding vowel forming the nucleus in the original CVCC syllable. This is also expected within the compound infinitive *fekr kærdæn*, meaning ‘to think’, where the same consonant is elided, at least in informal contexts. Naturally, neither of the above would be counted as an occurrence of a coda consonant cluster here.

This study is aimed at verifying to what extent Persian behaves in accordance with the universal patterns observed in languages of the world with respect to the sonority hierarchy. These observations have been brought about by typological studies such as Greenberg (1978) and Jany et al. (2007). It is hypothesized that while totally different data will serve as basis for investigation here than those referred to in studies considering words in isolation, the universal in question will still prove to hold in general.

**2. Review of Literature**

Kambuziya (2004) and Kambuziya and Zolfaghari Serish (2006) examine the validity of the sonority sequencing principle in Persian. The latter study is based on words extracted from Keshani’s and Moshiri’s dictionaries of Persian and classification of their syllables into two classes based on the vowels filling the nuclei. The classification is performed since the study is aimed to investigate a specific hypothesis about natural classes of Persian vowels, assumed by a number of researchers in the area. They find support for the hypothesis, observing that the principle holds for syllables with nuclei belonging to one class and not for those with nuclei belonging to the other. Kambuziya and Kheirabadi (2012) adopt a similar method to study the impact of syllable structure on the productivity of verbal derivational suffixes in Persian.

Rahbar (2012) provides a more theoretical than experimental discussion of the different behavior observed in terms of presence or absence of falling sonority in Persian VCC codas where the vowel belongs to either of two classes. While little explicit reference is made to any empirical data, she introduces Emami’s Persian dictionary as the inventory of Persian words she has checked to verify the state-of-affairs in Persian regarding certain phonological and morpho-phonological phenomena including this one.

In a computational study, Rahimi, Eslami, and Vazirnezhad (2012) investigate the syllable contact law in Persian, computing the marginal probabilities for each Persian consonant occurring at syllable boundaries. Their research is based on a Persian lexicon, called FLexicon, containing over 50,000 lexemes with phonemic transcriptions. The results demonstrate that falling sonority patterns occur more at syllable boundaries within the lexicon as well as the corpus, which is in line with SCL.

Rahimi, Vazirnezhad, and Eslami (2014) examine the distribution of sonority in the syllable structure of Persian. The results show that with some exceptions, Persian obeys SSP. They also observe that sonorant consonants occur more than expected no matter where in the syllable they are, and obstruents such as plosives occur as expected in pre-vocal contexts, which supports the idea that Perceptual Cue Salience plays an important role in shaping sonority sequencing in syllables. The results also provide support for Steriade’s Licensing by Cue hypothesis, which states that phonological contrasts are maintained in environments that provide better acoustic cues to the contrasts, and are neutralized in environments that provide poorer acoustic cues or no cues.

Mehraban (2010) examines the sonority sequencing principle in the Persian phonological system. Another research question she examines is whether phonological processes are performed as motivated by SSP. Having Examined 680 words with consonant clusters extracted from Moin’s dictionary of Persian, she finds that 68.67% of the words obey the principle, and that most of those that violate it are loan words, which sounds odd as long as a universal principle is concerned. What matters here regardless of the above claim in regard to borrowing is that she finds support for both hypotheses concerning the phonological system and processes.

Samare (2006) provides a detailed statistical analysis of Persian syllables also based on existing Persian words in isolation, reflecting type rather than token frequencies. His survey is rather descriptive, however. He does not make a mention of universal patterns, but describes Persian CVCC syllables in terms of each of their components.

The present study deals with Persian coda consonant clusters in the context of speech rather than isolated words, as it is in an environment of connected speech that analysis of segments and syllables in a phonological system makes perfect sense. This may not make much difference in the case of languages where the phonological word highly corresponds to the syntactic word. In Persian, however, this is not the case, which, from a pragmatic point of view, questions the validity of findings concerning phonological phenomena obtained from studies relying solely on data from lexicons.

**3. Methodology**

In this paper, the analysis was to be based on connected Persian speech, which most likely reflects Persian speakers’ choice of coda CC clusters, as the final way in which segments are grouped together and uttered, presumably quite different in many cases from how isolated words patterned individually but placed next to one another would behave. As promising as it seemed, the above-mentioned limitation imposed by the purpose of the research directed us toward a more difficult task than those assigned in similar earlier studies, concerned solely with examination of a single word at a time. In other words, it would not make sense if we had selected a static lexicon of Persian as the basis for our investigation, since that would obscure the reality of the syllables to be studied, leading us to results pointing to the state of affairs in a different domain.

An option would be to record instances of actual Persian speech. This would suffer two disadvantages, though. First, it would double the time and effort required without an advantage, as it would add the concern for proper data collection to that regarding the obligatory data analysis phase. Furthermore, typically lower error rates are expected in corpus-based studies relative to those involving manual collection of data, allowing for much interference from the experimental method that is taken.

Yet another possible solution would be to depend on a written text corpus of Persian as well as introspection to provide a well-sampled collection of pieces of connected Persian speech. This, however, would not eliminate the problems mentioned above in regard to recorded materials.

Therefore, we selected a speech corpus called FarsDat (http://www.rcisp.com/, Bijankhan et al., 1994) containing pieces of actual Persian speech, equipped with transcriptions and a range of annotations providing additional phonetic, phonological, and lexical information, part of which would turn out to be useful for our analytical purposes.

FarsDat was provided by the Research Center of Intelligent Signal Processing, established in 1992, operating in the form of six research groups in different areas of signal processing. FarsDat is a product of the Center’s group in charge of segment processing, one of Iran’s oldest research groups.

Inspired by the English counterpart, TIMIT, FarsDat contains a total of 6,080 phonetically-balanced sentences, uttered by speakers of different varieties spoken in Iran, not limited even to accents and dialects of Persian, but including varieties of other Iranian languages like Kurdish and Balochi. Since the present research was concentrated on the standard variety of Persian, i.e. the Tehrani dialect, however, use was made here only of the sentences uttered by the Tehrani speakers.

FarsDat audio files are accompanied by ones containing transcriptions of the contents, but which lack supra-segmental information including some on syllables. While the former files were listened to, the latter were further annotated with marks specifying syllable boundaries. The boundaries were determined by what was actually heard rather than supposed to be heard based on expatiations directed by pronunciations of isolated words. This explains the claim that the speech examined in the study is *actual*.

During the very first step of data collection, however, it was found that the extra difficulty involved in the search for CVCC syllables in connected speech had actually been underestimated, in such a way that even the whole collection of sentences uttered in the above corpus in Tehrani accent, Persian’s standard variety spoken in Iran’s capital, would not suffice to provide a sample of CVCC syllables large enough to be representative of the Persian language. The data collection procedure, therefore, was carried on with based on hours of speech randomly recorded from TV news programs. This was assumed to reduce the effects of the disadvantages mentioned above.

Subsequently, the CVCC syllables in the resulting annotated corpus were extracted. A total of 561 CVCC syllables were thus collected to be analyzed. Those that conformed to the sonority sequencing principle were listed separately from those that did not. Each list was assigned a Microsoft Office Excel worksheet. In either case, the qualities of all segments constituting the syllable were recorded as fields representing the onset, the nucleus, and the coda for (probable) future use. Each of the above worksheets, therefore, consisted of rows of records each containing, in turn, four columns. All the columns represented Persian consonants except the second, which hosted one of the six members of the Persian vowel inventory, /æ, e, o, ɑ, u, i/, as recognized by International Phonetic Association (1999).

In order to perform frequency analyses, we extracted reports from the Excel worksheets, based on the corresponding field values indicating the class of syllable codas each occurrence belonged to. The reports were included based on expectations formed by findings of previous research. In the next step, reports that appeared to suggest significant relashionship(s), whether confirming results of earlier studies or rejecting them or proposing totally novel ideas of existence of a certain pattern were specified. The reports were extracted from the databases through queries based on the last two columns, corresponding to the coda consonants.

**4. Results**

Each CVCC syllable encountered in the data was placed in one of the following groups based on the quality of its CC cluster’s components. Each section on heterogeneous sequences is concluded by a table summarizing the observed data with one instance from every sequence. Cells corresponding to unobserved sequences are left blank, which means that they are likely to get filled with instances from a larger amount of data.

**4.1. Codas in accordance with universal patterns**

4.1.1. Codas with a glide followed by a liquid

The only widely acknowledged modern-Persian glide is /j/, and the only Persian liquids are /l/ and /r/. Therefore, there are only two possibilities for the codas of this group: /jl/ and /jr/. As expected, there were all together only 8 instances in the data, including /mejl/ and /sejr/.

|  |  |  |
| --- | --- | --- |
| 1 2 | /l/ | /r/ |
| /j/ | /mejl/ | /sejr/ |

Table 1: Instances of glide-liquid sequences

4.1.2. Codas with a glide followed by a nasal

The same state of affairs as that mentioned about the previous group holds about this one, as just like in the case of liquids, there exist only a couple of nasals in Persian: /m/ and /n/. Here, as few as 4 instances were observed, including /bejn/ and /dejm/.

|  |  |  |
| --- | --- | --- |
| 1 2 | /m/ | /n/ |
| /j/ | /dejm/ | /bejn/ |

Table 2: Instances of glide-nasal sequences

4.1.3. Codas with a glide followed by an obstruent

It may be tempting to regard plosives as lower in sonority than affricates and also to consider them, in turn, still lower than fricatives. However, even though the class of sonorant segments is subcategorized into vowels, glides, liquids, and nasals in studies of the sonority hierarchy, it is by no means customary to do this for the opposing class of non-sonorant segments (that is, obstruent consonants), while in many languages, including Persian, the latter class contains the majority of the members of the phoneme inventory. Therefore, as many as eighteen obstruents, including eight plosives, eight fricatives, and two affricates were considered as involved in the present research. Nevertheless, Samare (2006) lists a nine-member subset of Persian obstruents that possibly ever occur in a CC coda beginning with /j/. There is no surprise, therefore, that only 19 instances were observed all throughout the data. Instances include /bejt/ and /sɑjz/.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 2 | /p/ | /b/ | /t/ | /d/ | /k/ | /g/ | /ɂ/ | /ʧ/ | /ʤ/ | /f/ | /v/ | /s/ | /z/ | /ʃ/ | /Ʒ/ | /x/ | /q/ | /h/ |
| /j/ |  | ɂejb | sɑjt | sejd | pejk |  |  |  |  | tejf |  | hejs | sɑjz |  |  | ʃejx |  |  |

Table 3: Instances of glide-obstruent sequences

4.1.4. Codas with a liquid followed by a nasal

Because these two sets each contain two members, as mentioned in 4.1.1 and 4.1.2, four sequences are possible: /ln/, /lm/, /rn/, and /rm/. Of these, /ln/ is not possible and /rn/ is extremely rare both due to the component phonemes’ same place of articulation (Samare, 2006). The data contained 5 occurrences of the other two sequences: /ɂelm/ and /gærm/.

|  |  |  |
| --- | --- | --- |
| 1 2 | /n/ | /m/ |
| /l/ |  | ɂelm |
| /r/ |  | gærm |

Table 4: Instances of liquid-nasal sequences

4.1.5. Codas with a liquid followed by an obstruent

Both /l/ and /r/ can be followed in a CC cluster by a variety of obstruents. The data exhibited a total of 76 occurrences of such sequences, including /zærd/ and /burs/.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 2 | /p/ | /b/ | /t/ | /d/ | /k/ | /g/ | /ɂ/ | /ʧ/ | /ʤ/ | /f/ | /v/ | /s/ | /z/ | /ʃ/ | /Ʒ/ | /x/ | /q/ | /h/ |
| /l/ |  | qælb |  | ʤeld | melk |  |  |  | dærʤ | zolf |  |  |  |  |  | tælx | xælq | solh |
| /r/ |  | ʧærb | kɑrt | særd | tork | gærg | ʃærɂ |  | mærʤ | bærf |  | dærs | værz |  |  | ʧærx | bærq | ʃærh |

Table 5: Instances of liquid-obstruent sequences

4.1.6. Codas with a nasal followed by an obstruent

The patterns observed here are pretty much similar to those seen in regard to the liquids in the same position. The data reflect the similarity with the number of such clusters reported as 83 instances, including /ʧænd/ and /ʃæms/.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 2 | /p/ | /b/ | /t/ | /d/ | /k/ | /g/ | /ɂ/ | /ʧ/ | /ʤ/ | /f/ | /v/ | /s/ | /z/ | /ʃ/ | /Ʒ/ | /x/ | /q/ | /h/ |
| /n/ |  |  |  | ɂænd | tɑnk | sæng |  |  |  | senf |  |  |  |  |  |  |  |  |
| /m/ |  | bomb | sæmt | ɂæmd |  |  |  |  |  |  |  | ʃæms |  |  |  |  |  |  |

Table 6: Instances of nasal-obstruent sequences

**4.2. Codas with a homogeneous sequence**

A CC cluster composed of two glides, which in Persian means a /jj/ sequence strictly, is certainly unattested in the position in question, as confirmed by Mahootian (1997), since it is possible only in a case of gemination, which occurs at the syllable boundary by definition. This also holds for other such sequences. The Persian liquid clusters composed of different segments, that is /lr/ and /rl/, are both impossible (Samare, 2006; Mahootian, 1997), which seems to be a language-specific restriction. In the case of the Persian nasals, unlike /nm/, /mn/ is possible (ibid.). Things are naturally different for obstruents. While the same consonant cannot be geminated within the coda as usual, as many as 217 instances of non-equal obstruent sequences were observed, which is not surprising, as long as they make up the most populated group of consonants in this analysis. Instances include /zeʃk/ and /mæqz/.

**4.3. Codas in contradiction with universal patterns**

4.3.1. Codas with a liquid followed by a glide

Based on Samare (2006), there are only four consonants in the Persian phoneme inventory that can ever follow the only Persian glide, /j/, and they are all categorized here as obstruents. The data confirm that, with no CVCC syllables found with the above requirement.

4.3.2. Codas with a nasal followed by a glide

Similarly to the case of the liquids, no trace was found of either of the Persian nasals preceding /j/ in a syllable coda cluster, as confirmed by Mahootian (1997).

4.3.3. Codas with an obstruent followed by a glide

As stated above, Samare (2006) lists four obstruents that can precede /j/. These include /h/, /ʃ/, /f/, and /ʔ/. Although it is controversial whether laryngeal segments like /h/ and /ʔ/ should be categorized as obstruents, and even as [- son] in the first place (Gussenhoven and Jacobs, 2011), they have been characterized so here. The only instance of an obstruent preceding a glide as the first member of a coda consonant cluster in the data was of /ʔ/, having occurred twice in exactly the same context: /sæɂj/[[1]](#footnote-1). Mahootian (1997) also considers this as a gap, most probably as a case of rarity.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 1 | /p/ | /b/ | /t/ | /d/ | /k/ | /g/ | /ɂ/ | /ʧ/ | /ʤ/ | /f/ | /v/ | /s/ | /z/ | /ʃ/ | /Ʒ/ | /x/ | /q/ | /h/ |
| /j/ |  |  |  |  |  |  | sæɂj |  |  |  |  |  |  |  |  |  |  |  |

Table 7: Instances of obstruent-glide sequences

4.3.4. Codas with a nasal followed by a liquid

Samare (2006) asserts that /n/ never accompanies either /r/ or /l/ due to their identical place of articulation. This is confirmed by the data, involving no instance of the segment /n/ preceding a Persian liquid. As far as /m/ is concerned, however, 7 occurrences were observed of it being the first consonant of a cluster ending in a liquid, including /hæml/ and /ɂomr/.

|  |  |  |
| --- | --- | --- |
| 2 1 | /n/ | /m/ |
| /l/ |  | hæml |
| /r/ |  | tæmr |

Table 8: Instances of nasal-liquid sequences

4.3.5. Codas with an obstruent followed by a liquid

Both segments /r/ and /l/ can follow different obstruent consonants within the coda of a syllable. The data displayed 67 total occurrences, including /ʤæɂl/ and /zohr/.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 1 | /p/ | /b/ | /t/ | /d/ | /k/ | /g/ | /ɂ/ | /ʧ/ | /ʤ/ | /f/ | /v/ | /s/ | /z/ | /ʃ/ | /Ʒ/ | /x/ | /q/ | /h/ |
| /l/ |  | tæbl |  | ɂædl | ʃekl |  | ʤæɂl |  |  |  |  | fæsl |  |  |  |  | ɂæql | ʤæhl |
| /r/ |  | sæbr |  | bædr | ʃokr |  | ʃeɂr |  | zæʤr | sefr |  | mesr | bæzr | qeʃr |  |  |  | mehr |

Table 9: Instances of obstruent-liquid sequences

4.3.6. Codas with an obstruent followed by a nasal

Along the same lines as observed in 4.3.5, both /m/ and /n/ are frequently preceded by a range of obstruents within Persian CC codas. 75 occurrences of the phenomenon were encountered in the data, including /qeʃm/ and /rokn/.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 1 | /p/ | /b/ | /t/ | /d/ | /k/ | /g/ | /ɂ/ | /ʧ/ | /ʤ/ | /f/ | /v/ | /s/ | /z/ | /ʃ/ | /Ʒ/ | /x/ | /q/ | /h/ |
| /n/ |  |  | mætn | ɂædn | rokn |  | ʃæɂn |  |  | dæfn |  | hosn | væzn | ʤæʃn |  |  |  | pæhn |
| /m/ |  |  | hætm |  | hokm |  | tæɂm |  | hæʤm |  |  | ʤesm | ɂæzm | xæʃm |  | toxm |  | sæhm |

Table 10: Instances of obstruent-nasal sequences

**5. Discussion**

As observed in the previous section, as many as 217 out of the 561 total instances of Persian CVCC syllables enumerated in the present research contained codas made up of consonants with the same position on the sonority hierarchy. In other words, about 38.6 percent of the codas neither obeyed nor violated universal patterns. This is not contrary to expectations, as statistics also suggests that the possibility of two members of a majority subset of segments filling both slots in a CC formula makes up enough permutations to allow the number of the cases of identical Cs to be so high despite the lack of (at least significant) sonority differences.

Adding up all the frequencies reported about different syllable compositions in favor of universal patterns yields a total of 195 (8 + 4 + 19 + 5 + 76 + 83). The same calculation made for the syllables contrary to universal patterns obtains 149. The above numbers mean around 34.8 and 26.6 percent, respectively, of all possible syllables. Ignoring the syllables discussed in the previous paragraph would give about 56.7 and 43.3 percent of all syllables for which arguments about universal patterns being obeyed or violated can be made (which means 344 instances out of the total 561). While the above 13.4 percent difference cannot be neglected, it does not provide a great support for the tendency claimed to hold universally either. This will be addressed further in the next section along with a consideration of a possible bias.

A comparison of the coda patterns based on specific orders of sonority values also turns out to be fruitful. While the liquid-glide order is prohibited, there were 8 occurrences of the opposite order. Whereas the nasal-glide order is not attested, the opposite order was represented by 4 instances. While there were only a couple of occurrences of the obstruent-glide order, this number was 19 for the opposite order. The unexpected nasal-liquid order exhibited 7 instances, while the opposite expected order equaled 5. There were 67 occurrences of the obstruent-liquid order, while the opposite order had 76 representatives. Similarly, the obstruent-nasal order was observed in 75 instances, a little fewer than the 83 occurrences identified as displaying the opposite order.

As seen above, it is only in the case of the nasal-liquid order that the cases contradicting universal patterns outnumber those contributing to their validity. While the above difference does not appear significant, the same claim may be made about each of the other specific cases listed above, except, probably, for that of the obstruent-glide order (that is, that including the extremes of the sonority hierarchy). Nonetheless, a comparison between the sums of the above numbers, as made earlier in this section, does reveal an overall tendency toward accordance with universal patterns.

**6. Conclusion**

The results of the present study performed on Persian provide a rather tentative confirmation of the ideas suggested through the sonority sequencing principle and the closely related syllable contact law about universal patterns of formation of consonant clusters within syllables in the world’s languages. Certain limitations in the phonological system of Persian as well as particular phonotactic constraints imposed by the suprasegmental phonology of the language may be argued to cast doubt on the possible conclusions to be drawn from the findings of this investigation. These are both best illustrated through the case of the Persian sole glide, /j/. First, it is considered the only one of its kind, reducing the number of the permutations available to the phonology to only one. Second, the segment is absolutely incompatible with all the liquids and nasals of the language, restricting the possibilities even further. All these facts, however, make up part of the overall tendency toward universal patterns, and therefore cannot be considered as obscuring the analysis of the data.

For the reasons mentioned earlier—that is, processes leading to elimination of CVCC syllables—an extensive amount of data collection is required, making the procedure a time-consuming one. Two solutions can be considered for improvement of similar research. The more accessible one is to employ more and more phonologists or trained volunteers to participate in the data collection procedure. The other solution is to initially develop a fully annotated (i.e., tagged with additional pieces of information) corpus of Persian speech devised specifically for syllable structure analysis purposes. In addition to the assistance it provides for the present research to be conducted, it can be applied to probable replications of this research with larger amounts of data, perhaps also more dependable.

**7. References**

Bijankhan, M., Sheikhzadegan, J., Roohani, M. R., Samare, Y., Lucas, C., and Tebyani, M. (1994). FARSDAT-The speech database of Farsi spoken language. Proceedings of the Australian Conference on Speech Science and Technology. Australia: Melbourne.

Greenberg, J. H. (1978). *Universals of human language, vol. 2: Phonology*. Stanford: Stanford University.

Gussenhoven, C., and Jacobs, H. (2011). *Understanding phonology*. 3rd ed. London: Hodder Arnold.

International Phonetic Association. (1999). *Handbook of the international phonetic association: A guide to the use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Jany, C., Gordon, M., Nash, C. M., and Takara, N. (August, 2007). How universal is the sonority hierarchy?: A cross-linguistic acoustic study. **Proceedings of the 16th International Congress of Phonetic Sciences.** Germany: Saarbrucken.

Kambuziya, A. (2004). Tasire vâkeye haste bar gozineše hamxânhâye xušeye pâyane [Effects of nucleus vowel on selection of coda cluster consonants]. Pažuhešnâmeye adabe qenâyi [Journal of Lyrical Literature Researches], 2, 105-14 (in Persian).

Kambuziya, A., and Kheirabadi, R. (2012). The impact of syllable structure on the productivity of verbal derivational suffixes in Persian. Language Related Research, 3 (11), 143-56.

Kambuziya, A., and Zolfaghari Serish, M. (2006). Sonority hierarchy principle in cvcc syllable of Persian. The International Journal of Humanities, 13 (1), 107-22. Tehran: Tarbiat Modarres University.

Mahootian, S. (1997). *Persian*. New York: Routledge.

Mehraban, M. (2010). *Barrasiye osule rasayi dar nezâme vâjiye zabâne fârsi* [Investigation of sonority principles in Persian phonological system] (PhD. Dissertation). Mashhad: Ferdowsi University (in Persian).

Rahbar, E. R. (2012). Aspects of Persian phonology and morpho-phonology (Doctoral dissertation, University of Toronto).

Rahimi, A., Eslami, M., and Vazirnezhad, B. (July, 2012). Syllable contact law in Persian. Proceedings of the 2nd Iranian Conference on Computational Linguistics. Iran: Tehran.

Rahimi, A., Vazirnezhad, B., and Eslami, M. (2014). Asle tavâliye rasâyi dar zabâne fârsi [The sonority sequencing principle in Persian]. Pardâzeše alâem va dâdaeha [International journal of signal and data processing], 11 (1), 87-94. Tehran: Research center of intelligent signal processing (in Persian).

Samare, Y. (2006). *Âvâšenâsiye zabâne fârsi: Âvâhâ va sâxte âvâiye* *hejâhâ* [*Phonology of the Persian language: Sounds and the sound pattern of syllables*] (2nd ed.). Tehran: Iran University Press (in Persian).

Selkirk, E. (1984). On the major class features and syllable theory, in Aronoff and Oehrle (Eds.) *Language sound structure: Studies in phonology.* Cambridge: MIT Press.

1. A prescriptive point of view may suggest that such strings are actually implausible to occur in Persian speech. Thanks to empirical data, hypotheses like that, sometimes based on spontaneous intuition or overgeneralization, can be examined, and each may be retained or rejected, as observed throughout this section. [↑](#footnote-ref-1)