7. Summary

This thesis is a statistical description of the properties of the syllable in Modern Standard Persian (Tehran dialect), with special attention to phonotactic constraints. The purpose is to describe what kinds of syllables are permitted and which are favoured in the language. One central theme in the analysis is the hypothesis that sonority decreases towards the syllable margins. This hypothesis is tested against the data of this study. Another topic treated in the thesis is the etymological composition of the data. A substantial portion of the vocabulary in Modern Persian comes from foreign sources, and this raises the question of whether or not the data of foreign origin differs from that of Persian origin with respect to the properties of the syllable, and, if so, how? Are the possible differences minor, or do they imply that foreign elements have changed the nature of Persian syllables in any crucial way?

The syllable in this study is assumed to have a hierarchical structure, consisting of the onset and the rhyme, which in turn is divided into the nucleus and the coda. What motivates the assumption of the hierarchical structure is the hypothesis that the vowel and the following consonant(s) in the syllable form some kind of a unit, whereas the onset functions independently of the rhyme.

The Persian syllable system is very simple. There are only three types of syllables: CV, CVC and CVCC. In other words, the onset is always filled, but no complex onsets (of two or more consonants) are allowed. The coda can be zero, or it may be filled by one or two consonants. That is, consonant clusters are permitted only in the coda. The nucleus is always occupied by one vowel.

The phonotactics in this study is concerned with the following issues: Which segments are permitted to occur in which slots (onset, nucleus, coda)? What consonant clusters can (and can not) occur in the coda? Are there any restrictions on the combination of onset+nucleus and nucleus+coda? Restrictions

on onset+nucleus sequences would argue against the assumption of a hierarchical structure of the syllable.

Sonority, being phonetically closely connected with the degree of blockage of the airstream, forms a continuum where one, the least sonorous end, is represented by voiceless stops, and the other, most sonorous end, is represented by open vowels. The rest of the segment classes come between these extreme ends, i.e. fricatives, nasals, liquids, glides, close vowels, and mid vowels. Sonority is thought to be an important factor in the organisation of segments in the syllable. The hypothesis is that sonority decreases when we move from the nucleus towards the syllable edge. Sonority in this study is considered in connection with what is permitted, but also with what is favoured. The main questions are as follows:

(1) Does the hypothesis of decreasing sonority predict which coda clusters can and can not occur in Persian? In case there are counter-examples, do they differ in terms of frequencies from those clusters that conform to the hypothesis?
(2) Are open, mid and close vowels in the data equally frequent or not? In other words, is there a correlation between the sonority of the vowel, and its frequency?

The basic unit of analysis in this study is the syllable. The set of syllables are obtained from 10175 dictionary words constituting the Data of Words (DW). The DW contains 24135 syllable tokens, but out of these only 2701 are different. They form the Data of Basic Syllables (DBS), and they are the primary object of this study.

The DBS contains 137 CV syllables, 1792 CVC syllables, and 772 CVCC syllables. The percentages of the different syllable types in the DBS and in the DW are as follows:

Table 7.1: *The percentages of different syllable types in the DBS and the DW*

Syllable types	DBS	DW
CV	5%	46%
CVC	66%	49%
CVCC	29%	5%
Total	100%	100%

The percentages show that CV syllable tokens cover nearly half (46%) of the DW syllables, but only 5% of the basic syllables in the DBS. This means that basic CV syllables are among the most frequent ones. Their total number is small, only 137. The most common syllable type in both data is CVC with two-thirds of the DBS and almost half (49%) of the DW syllables being of this type. The percentages imply that basic CVC syllables are found across the whole frequency continuum. Finally, the CVCC syllables cover only 5% of the DW, but almost 30% of the basic syllables in the DBS. This suggests that the type CVCC contains mainly low frequency syllables.

The importance of the CV type in Modern Persian is also reflected by the fact that practically all (over 99%) of the possible CV sequences permitted by the Persian phoneme system are actually found in the data. The corresponding percentage for CVC syllables is considerably lower, 56%, and for CVCC syllables as low as 1%.

Words in Modern Persian come from various etymological sources. Arabic is the biggest contributor of words, French comes second. The DW contains words from 13 different languages; some words have elements from two sources. The DBS is divided into two subdata: (basic) syllables coming from words of Persian origin (DPS), and (basic) syllables that are found only in words of non-Persian origin (DNPS). The proportions of syllable types somewhat differ in different data. Tables 7.2-3 show these differences:

Table 7.2: *The percentages of the Persian and non-Persian words in the DW*

Source	Data of Words (DW)
Persian	39%
non-Persian	60%
mixed	1%
Total	100%

Table 7.3: *The percentages of Persian and non-Persian basic syllables and proportions of different syllable types in the DPS and DNPS*

Source	Data of Basic Syllables (DBS)	Syllable types		
		CV	CVC	CVCC
Persian	51%	96%	56%	35%
non-Persian	49%	4%	44%	65%
Total	100%	100%	100%	100%

Roughly 40% of the words in the DW are of Persian origin and 60% of foreign origin. In each group, Arabic is clearly the biggest foreign contributor. It has a proportion of 53% of the words in the DW. Basic syllables are nearly half of Persian and half of foreign origin. Again, Arabic has a big proportion, 40% of the DBS. Of the syllables types, nearly all (96%) of basic CV syllables are found in words of Persian origin, but the percentage decreases when the complexity of the syllable increases. Somewhat more than half (56%) of CVC syllables are of Persian origin. The proportion of Arabic in this syllable type is 38%. The proportion of Persian is lowest in the CVCC syllable, only 35%, while the Arabic proportion is 53%.

The profiles of consonant frequencies in the DW and the DBS do not differ dramatically from each other (Figures 2.2, 3.2 and 3.3). The four least frequent consonants, /g/, /p/, /č/, /ž/, in descending order, are the same in both data. They are all consonants that are missing in Arabic, the biggest foreign source. The most frequent consonant is the same in both data, namely, /r/. The three most frequent consonants in the DW are all the sonorants, i.e. /r/ and the nasals. Their proportions in the DBS are lower, which means that they occur more in higher frequency syllables. As to the proportions of manner subclasses,

the biggest differences between the data are in the proportions of fricatives and nasals (Figure 3.5). The DW has a bigger proportion of fricatives than the DBS (in the DW 25%, in the DBS 31%), while the situation is the opposite for nasals (17% in the DW, 12% in the DBS). This means that higher frequency syllables favour nasals, while there are a great number of fricatives also in low frequency syllables. The proportions of dentals also differ (Figure 3.7): higher frequency syllables favour dentals, while palatals and uvulars concentrate more in lower frequency syllables.

Vowel frequencies in the DW correlate positively with the degree of the sonority of the vowel (Figure 2.5); open vowels have the highest frequencies, close vowels have the lowest frequencies. Moreover, each front vowel is more frequent than its back pair. In the DBS, the situation is not quite the same (Figure 3.8) since the differences between the percentages of vowels are smaller than in the DW, and the proportion of /a/ in the DBS is in the same range as the proportions of mid vowels rather than the other open vowels. This means that higher frequency syllables favour more sonorous vowels. In the DBS, vowels seem to be divided into two series, the short vowels /ä, e, o/, which are more frequent than the long vowels /a, i, u/. In both series, sonority correlates positively with vowel frequency.

The syllable type CV does not have any (systematic) constraints either on the onset consonant or the nuclear vowel. Out of the 138 theoretically possible CV sequences in the language, all but one, /žä/, are found in the data. That is, more than 99% of the possible CV sequences are used in the language.

Of the theoretically possible CVC sequences, a little over half (56%), 1792 syllables, are found in the data. The onset and the coda are somewhat asymmetric: the proportion of non-continuants (stops and affricates) in the onset (42%) is bigger than in the coda (36%), i.e. the coda favours continuants (Figures 4.3 and 4.6). The proportion of dentals is also bigger in the coda than in the onset (43% of coda consonants, 38% of onset consonants).

There are some differences between the DPS and the DNPS in onset consonant frequencies (Figures 4.13, 4.16, 4.17 and 4.19), where /p/, /g/ and /č/ have remarkably low frequencies in the DNPS, but not in the DPS, while the foreign proportion is conspicuous in /G/, /ʔ/ and /f/. The DPS has a big

proportion, 59%, in stops of the DBS, while the foreign proportion is big in fricatives, nasals (54% in both), and in the glide (63%).

Special features of the nuclear vowels of CVC syllables in the DBS are, first, that their proportions differ rather little from each other, and second, that the proportion of /i/ is in the same range as the proportions of the mid vowels. If short and long vowels are treated as separate sets, then each set shows a clear positive correlation between vowel frequency and sonority.

The two etymological data, the DPS and the DNPS, differ with respect to the proportions of nuclear vowels (Figures 4.29-32). Open vowels have clearly the biggest proportion in the DPS, and /a/ has the highest percentage, 25%. Mid vowels have the lowest percentages, which, however, are near to that of close, vowels. If long and short vowels are treated as separate series, then each series has a positive correlation between vowel frequency and sonority. A characteristic of the CVC syllables in the DPS is that long vowels together have a bigger proportion than short vowels (53% long V, 47% short V). In the DNPS, the biggest proportions belong to the mid vowels, while the proportions of open and close vowels are nearly the same. Of the open vowels, /ä/ is more frequent than /a/. Vowel frequency and sonority do not correlate either in the whole set of vowels or in the separate series of long and short vowels in the DNPS. These partly opposite properties of vowels in the DPS and the DNPS bring about the situation exhibited by the DBS, in which frequency differences between both individual vowels and vowel classes are very small.

The syllable type CVCC in the DBS contains 772 syllables, which is only 1% of all the theoretically possible sequences. The proportions of manner subclasses in the onset of CVCC syllables are very similar to those of CVC syllables: the biggest difference is in the proportions of the glide (4% in the DBS, 1% in the CVCC). This means that the syllable type does not affect the onset. Percentages of manner subclasses in the onset and the coda of CVC and CVCC syllables in the DBS is shown in Tables 7.4-A and 7.4-B. These tables are based on Figures 4.3 and 5.2:

Table 7.4-A: *The percentages of the manner subclasses in the onset of the CVC and CVCC syllables in the DBS*

Onset	Stops	Affricates	Fricatives	Nasals	Liquids	Glide
CVC	36.16%	6.19%	31.36%	10.21%	12.00	4.07%
CVCC	37.14%	5.58%	32.99%	11.95%	11.04%	1.30%

Table 7.4-B: *The percentages of the manner subclasses in the coda of the CVC and CVCC syllables in the DBS*

Coda		Stops	Affricates	Fricatives	Nasals	Liquids	Glide
CVC		32.37%	3.85%	33.54%	12.56%	13.84%	3.85%
CVCC	C_1	19.82%	1.94%	34.33%	14.51%	24.48%	4.92%
	C_2	45.21%	2.72%	22.15%	10.88%	18.13%	0.91%

The two coda positions of the CVCC syllable differ from each other (Figures 5.5 and 5.8). As the above table shows, the first position, C_1, favours continuants (fricatives and sonorants), the second position, C_2, favours stops. The table also shows that the coda of CVC syllables differs both from the C_1 and C_2 of CVCC syllables. In other words, the syllable type has an effect on the properties of the coda.

Nuclear vowels in CVCC syllables form two clearly separate series of short and long vowels. Short vowels cover 87% of vowel tokens; /ä/ alone has a proportion of 51%. The proportion of long vowels is only 13%; the most frequent long vowel, /a/, has a proportion of 8%. Thus, sonority and frequency correlate positively in both series. The reason for the strong preference of short vowels can be in the syllable weight, i.e. long nuclear vowels in CVCC syllables produce extra heavy syllables, which is not optimal.

Vowels in CVC and CVCC syllables of the DBS show positive correlation between vowel frequency and sonority; in CV syllables all vowels are practically equally frequent. The positive correlation shows up even more clearly and in all syllable types if we take the 50 most frequent syllables of each type in the DBS and count vowel frequencies. The results are given in Table 7.5 (next page).

Table 7.5: *Frequencies of the vowels of 50 most frequent syllables of each syllable type in the DBS*

Vowels	Frequency		
	CV	CVC	CVCC
ä	14	29	36
a	18	14	2
e	12	5	4
o	2	2	6
i	4	0	0
u	0	0	2
Total	50	50	50

In all syllable types, the most sonorous (open) vowels are the most frequent ones, and their proportion increases when the complexity of the syllable increases. In the CVCC type, only the short open vowel has a high frequency.

Out of the 772 coda consonant clusters of CVCC syllables, 209 different combinations are found. Two tendencies can be discerned in the nature of the clusters, a tendency towards the dissimilarity of cluster members and decreasing sonority at the syllable edge (i.e. the C_2 consonant is less sonorous than the C_1 consonant). The tendency towards dissimilarity is seen in the fact that consonants from the same manner subclass cluster less often than consonants from different subclasses. And even when there is a fair amount of clustering, as is the case of stops and fricatives, where the number of existing clusters is nearly 30% of the theoretical maximum, cluster frequencies tend to be low. Likewise, consonants from the same place class tend not to cluster. The only exception is dentals, where the proportion of existing clusters out of the theoretical maximum is 55%. Moreover, five out of the ten most frequent CC clusters are combinations of two dentals.

The tendency towards decreasing sonority at the syllable edge is seen most clearly in frequencies of cluster tokens, but also in the number of clusters, as the Table 7.6, which is based on Table 5.40, shows.

Table 7.6: *Number of different CC clusters, cluster tokens and their percentages in different cluster types of the DBS*

Cluster types	Clusters	%	Cluster tokens	%
Decreasing sonority	93	45	468	60
Increasing sonority	78	37	235	30
Same sonority	38	18	69	10
Total	209	100	772	100

Table 7.6 indicates that the least favoured combination, both in terms of the number of clusters and especially with respect to the number of cluster tokens, is the type where both members come from the same sonority level. The type where sonority decreases at the syllable edge is the most favoured one, and very clearly so in terms of the number of cluster tokens. However, some clusters with increasing sonority (e.g. *stop+liquid* and *fricative+liquid*) have both a fair number of clusters and cluster tokens. It seems thus that the tendency towards dissimilarity is stronger than the tendency of decreasing sonority.

CC clusters show evidence of phonotactic constraints. Certain segments are not permitted in certain slots, e.g. /p/, /č/, and /ž/ do not occur in C_1. In addition, certain kinds of sequences are not allowed. Some of the constraints can be seen as motivated by the tendencies discussed above. Thus, the tendency towards dissimilarity of cluster members – and in some cases also the tendency towards decreasing sonority – can account for the following constraints:

(1) Clusters of identical consonants are not permitted.
(2) Clusters where the members have the same place and manner of articulation and differ only in voicing are not permitted.
(3) Clusters consisting of sibilants or of liquids are not permitted.
(4) Clusters consisting of back area (velar, uvular, glottal) stops, or back area fricatives are not permitted.

In addition to these absolute prohibitions, there is a strong tendency for the following:

(5) Back area stops and fricatives tend not to cluster with each other.

The majority (65%) of CVCC syllables in the DBS come from words of foreign origin. Both the DPS and the DNPS show similar properties of the nuclear vowel, with /ä/ having a proportion of roughly half of all tokens, and short vowels being more frequent than long vowels. The difference between the data is that the proportion of back vowels in the DPS is bigger than in the DNPS. In the onset, the foreign data has a somewhat bigger proportion of fricatives and nasals, but a smaller proportion of stops and affricates than the DPS. In the coda, the two etymological data differ more, the DPS has almost exclusively continuants in the C_1 position, while the DNPS has also affricates and a high number of stops. In the C_2 position, 73% of the DPS consonants are non-continuants, i.e. stops or affricates, while only 35% of the DNPS consonants in C_2 are non-continuants, but as big a proportion as 39% in C_2 are sonorants. In other words, the DPS has a very strong tendency to place non-continuant consonants at syllable edges and to prefer continuants in C_1. This suggests that the two data also differ in the nature of CC clusters. The preferred cluster type in the DPS is one with decreasing sonority. The sequences *sonorant+stop/fricative* and *fricative+stop* have the highest number of clusters. There is only one type with decreasing sonority that has several clusters in the DPS, namely, *fricative+sonorant*. The DPS has only a few clusters where the members come from the same manner subclass. Foreign sources have contributed new clusters to all three types, but their proportions are biggest in clusters with increasing sonority and in clusters with members from the same manner subclass. Thus, loanwords have not changed the basic pattern of CC clusters, but they have added greatly to the number of those clusters where the principle of decreasing sonority does not work.