4. CV and CVC syllables

4.1. CV syllables

It was mentioned in 3.1 that there are 138 (= 23x6) theoretically possible different CV syllables in Modern Persian. All but one of them are found in the DBS. The missing syllable is /žä/. This gap is purely accidental. There are no systematic restrictions on the combinations of onset and nucleus in CV syllables.[40]

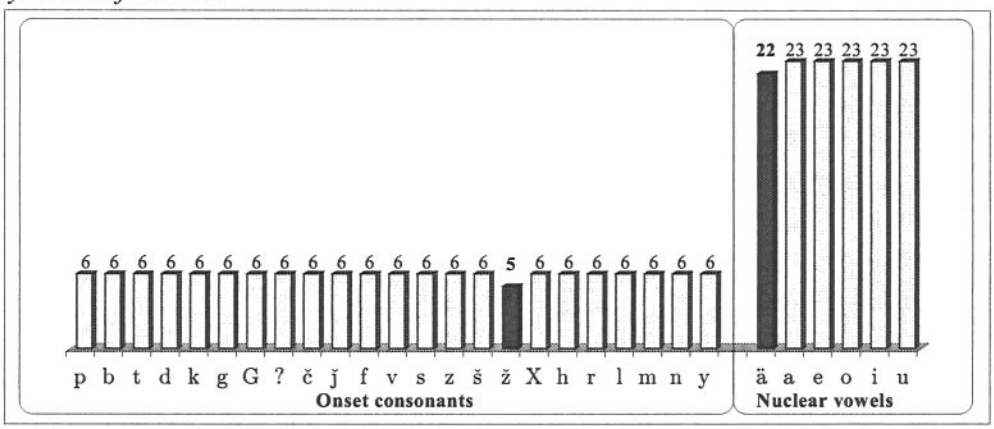
Figure 4.1 shows the frequencies of consonants and vowels in the onset and nucleus of CV syllables. The diagram is quite monotonous, since there is just one gap both in consonants and in vowels: the frequency of each consonant is 6, except for /ž/, whose frequency is 5. Similarly, the frequency of each vowel is 23, except for /ä/, whose frequency is 22.

Etymologically Persian syllables contain 132 syllables out of the 137 in the DBS. The five syllables missing in the DPS are /ʔe/, /fu/, /Gi/, /vo/ and /žo/. The syllables /ʔe/ and /fu/ are found both among etymologically Arabic and French syllables, /Gi/ is found among etymologically Arabic, Turkish and Greek syllables, /vo/ is found only among etymologically Arabic syllables and /žo/ is found exclusively among etymologically French syllables.

---

40 The missing syllable /žä/ is found in highly archaic words which are not used in Modern Persian. Some examples of these kinds of words are /žä-da-Ga/ 'idolater, idol worshipper', /žä-Gäng/ 'hiccup', /žä-ra-Gäng/ 'sandy land', /žä-Ga-re/ 'millet bread', /žä-fi-de/ 'wet', and /žä-Gav/ 'prostitute'. As referred to in Chapter 1, the data used for this study has excluded highly archaic words. Examples of archaic words cited in this study come from the following sources: Dehkhoda (1993), Padeshah (1983), Borhan-e Tabrizi (1963) and Nafisi (1976).

Figure 4.1: *Frequencies of the consonants and vowels in the onset and nucleus of CV syllables of the DBS*



4.2. CVC syllables

4.2.1. CVC syllables in the DBS

As was mentioned in 3.1, the number of theoretically possible CVC syllables in Persian is 3174 (= 23x6x23). Out of these, 1792 syllables are found in the DBS, which is a little more than a half (56%). There is a clear increase in gaps when the syllable complexity increases; the proportion of gaps is less than 1% in CV syllables, while it is 44% for CVC syllables.[41] In the following, I shall deal with segment frequencies in different positions of the CVC syllable.

---

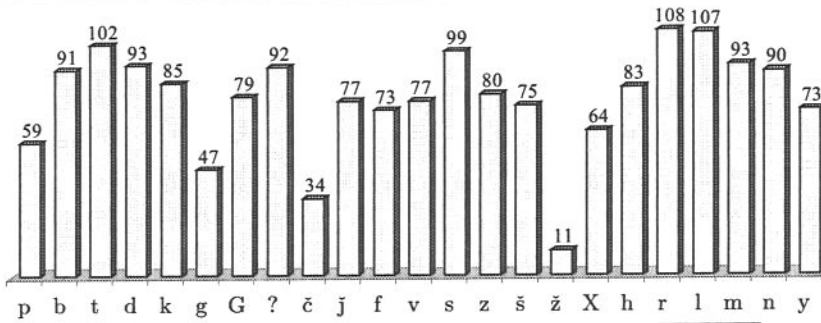
[41] A number of the missing syllables may be found in highly archaic words. For example: /šob/, /šäp/, /šoǰ/, /šäš/, /beǰ/, /bäǰ/, /boč/, /boχ/, /bäž/, /bož/, /čok/ (in the word /šä-čok/), /čač/ (in the word /čač-le/), /čaχ/ (in the word /čaχ-šuk/), /čäb/ (in the word /čäb-čä-le/), /čob/ (in the word /čob-čor-ge/), /doǰ/, /duž/, /gop/, and /got/ (in the word /got-rom/), /goǰ/. Some of the missing syllables can also be found in non-Tehrani dialects of Persian which have been excluded from the data. For example: /dak/, /däp/ and /dek/ (Shushtari dialect), /gät/ (Mazandarani dialect), /gäb/ (in some dialects of Gonabad, Khorasan and Gilan), /gäp/ (in Lori dialect and Tehrani slang). The absence of some of the syllables is due to the phonological changes that a number of phonemes have undergone "in certain periods of history" (Natel Khanlari 1968: 93). For example, (see Natel Khanlari 1987: 68-83):

(1) /v, p/ → /b/: /tav/ → /tab/, /pin/ (as in the word /zupin/) → /bin/, and /päd/ (as in the word /sä-päd/) → /bäd/
(2) /č, ž, g/ → /ǰ/: /ruč/ (as in the word /sa-ruč/) → /ruǰ/, /mač/ (as in the word /ʔa-mač/) → /maǰ/, /žän/ (as in the word /lä-žän/) → /ǰän/, /baž/ → /baǰ/, /zag/ → /zaǰ/

4.2.1.1. Onset

Figure 4.2 presents consonant frequencies in the onset of CVC syllables. This figure shows that all consonants can occur in the onset, but none of them occurs in all possible _VC combinations. For each consonant the frequency is less than the theoretical maximum 138 (= 6 x 23). The most frequent consonants, i.e. those with the fewest gaps, are the liquids /r/ and /l/, which have gaps in less than 25% of all the theoretical possibilities. The least frequent consonant is /ž/, with gaps in 92% of the theoretically possible combinations. A great majority of the consonants (18 out of 23) occur in more than half of the theoretically possible combinations.

Figure 4.2: *Frequencies and percentages of consonants in the onset of the CVC syllables of the DBS*



p	3.29%	d	5.19%	G	4.41%	j	4.30%	s	5.52%	ž	0.61%	r	6.03%	n	5.02%		
b	5.08%	k	4.74%	?	5.13%	f	4.07%	z	4.46%	χ	3.57%	l	5.97%	y	4.07%		
t	5.69%	g	2.62%	č	1.90%	v	4.30%	š	4.19%	h	4.63%	m	5.19%	100.00%			

DO: r>l>t>s>d,m>ʔ>b>n>k>h>z>G>v,j>š>f,y>χ>p>g>č>ž

(3) /ž/ → /z/: /žäχ/ (as in the word /du-žäχ/) → /zäχ/, /žaf/ (as in the word /gä-žaf/) → /zaf/, /žar/ (as in the word /kar-žar/) → /zar/, and /žän/ (as in the word /žän-gar/ → /zän/

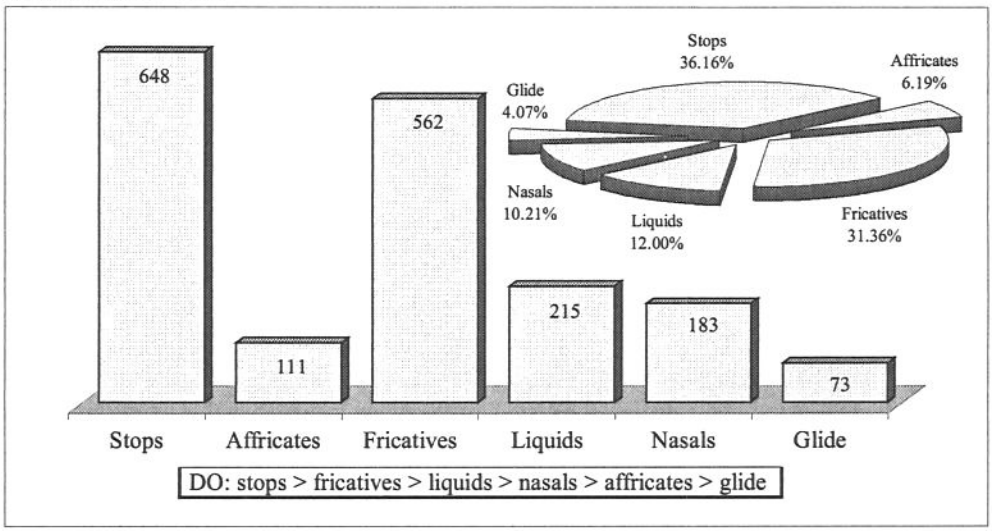
(4) /p/ → /f/: /pam/ → /fam/

(5) /b/ → /p/: /čab/ (as in the word /čab-lus/) → /čap/

Dealing with these kinds of changes needs and independent diachronic study, which is beyond the scope of the present research.

Figure 4.3 shows frequencies and percentages of onset consonants in different manner subclasses. The figure shows that the biggest groups are stops, with a proportion of 36%, and fricatives, with a proportion of 31%. The proportions of the rest range from 12% (liquids) to 4% (the glide).

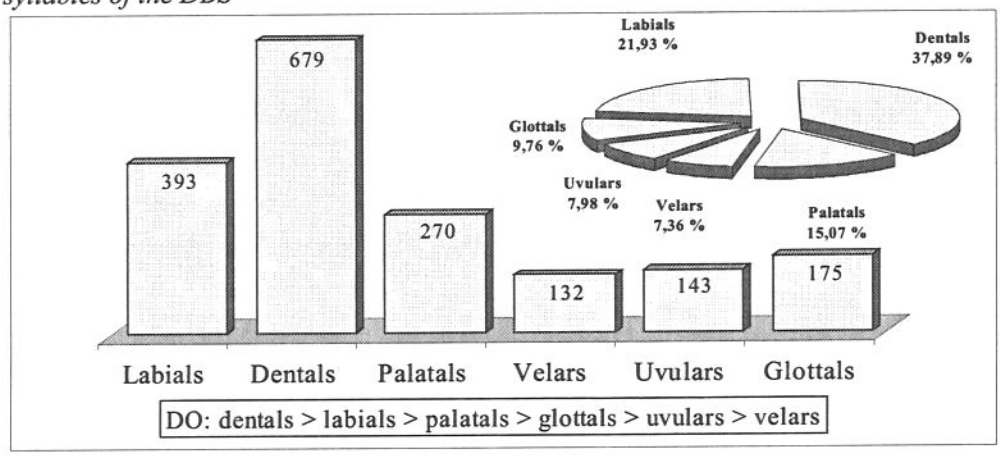
Figure 4.3: *Frequencies and percentages of manner subclasses in the onset of the CVC syllables of the DBS*



Of the onset consonants, nearly three quarters (74%) are obstruents and one quarter (26%) are sonorants. Of the obstruents, 58% are voiceless and 42% are voiced.

Figure 4.4 gives the frequencies and percentages of different place classes in the onset. As the figure shows, dentals are the most common class in the onset, with a proportion of 38%. Labials come next, with a proportion of 22%. The proportions of the rest range from 15% (palatals) to 7% (velars).

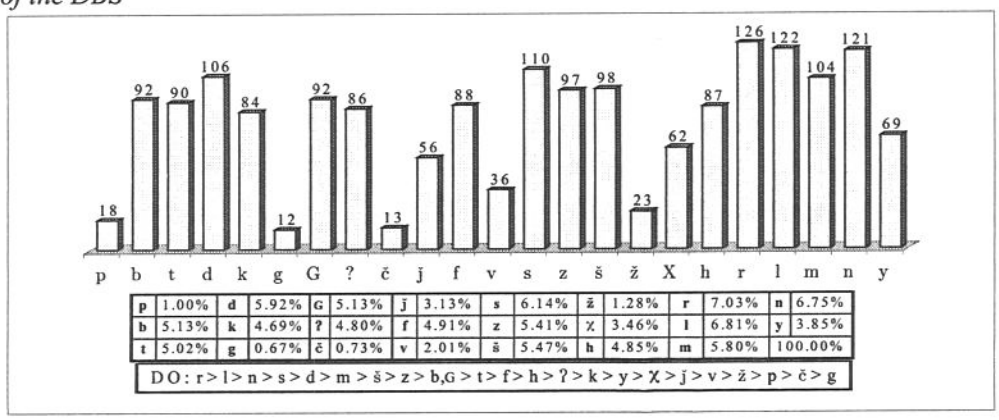
Figure 4.4: *Frequencies and percentages of the place classes in the onset of the CVC syllables of the DBS*



4.2.1.2. Coda

Figure 4.5 shows coda consonant frequencies of CVC syllables. As in the onset, all consonants can occur in the coda, but not in all combinations of CV_. The three most frequent consonants, /r, l, n/, are all sonorants, and they form 21% of the total number of coda consonants in the DBS. The percentage of gaps for /r/, the most frequent consonant, is a little below 10% of the maximal frequency of 138. On the other hand, there are several consonants with very low frequencies (i.e. with a high percentage of gaps) in the coda, e.g. /ž, p, č, g/.

Figure 4.5: *Frequencies and percentages of consonants in the coda of the CVC syllables of the DBS*



p	1.00%	d	5.92%	G	5.13%	j	3.13%	s	6.14%	ž	1.28%	r	7.03%	n	6.75%
b	5.13%	k	4.69%	?	4.80%	f	4.91%	z	5.41%	χ	3.46%	l	6.81%	y	3.85%
t	5.02%	g	0.67%	č	0.73%	v	2.01%	š	5.47%	h	4.85%	m	5.80%		100.00%

DO: r > l > n > s > d > m > š > z > b,G > t > f > h > ? > k > y > χ > j > v > ž > p > č > g

Obstruents have a proportion of 70% and sonorants 30% of the total number of the consonant tokens. Of the obstruents 59% are voiceless and 41% voiced.

Figure 4.6 presents consonant frequencies of different manner subclasses in the coda. The figure shows that there are two big classes: fricatives, with a proportion of 34%, and stops, with a proportion of 32%. The percentages of the other classes range from 14% (liquids) to 4% (affricates and the glide each).

Figure 4.6: *Frequencies and percentages of manner subclasses in the coda of the CVC syllables of the DBS*

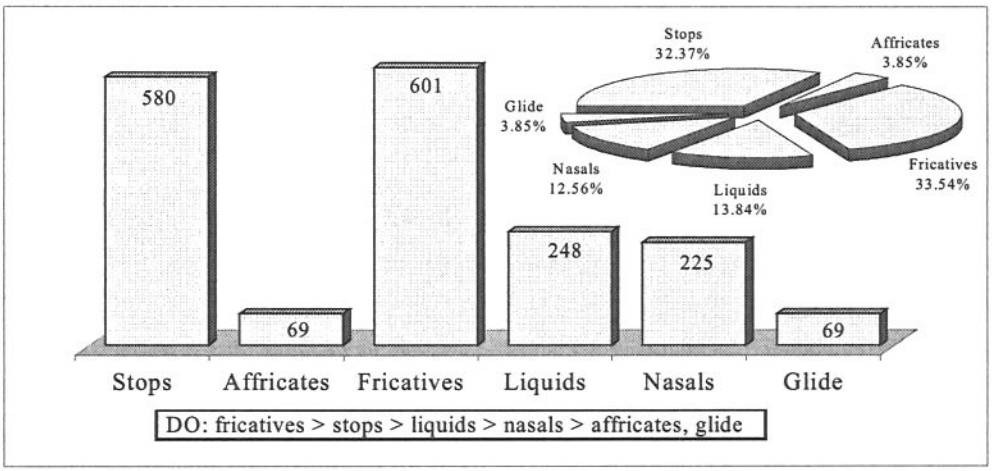
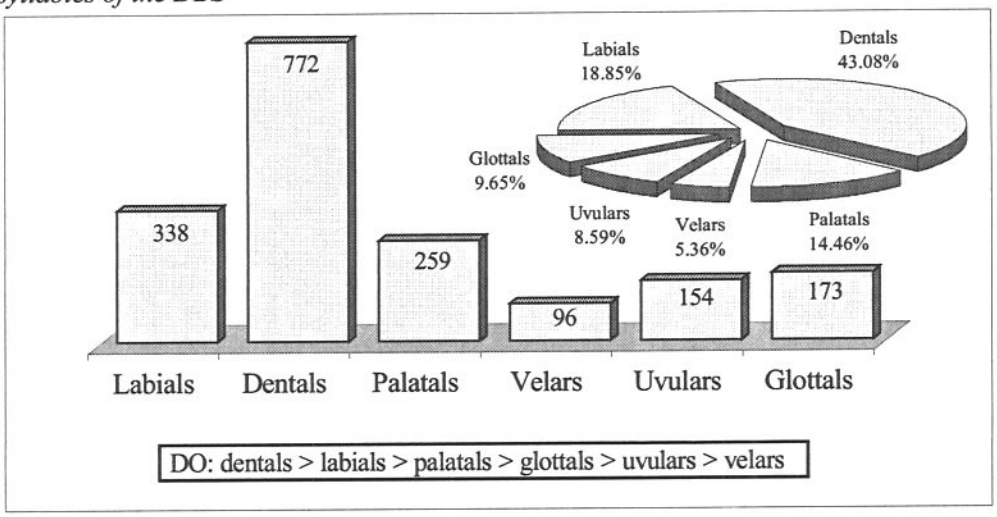


Figure 4.7 gives frequencies and percentages of different place classes in the coda. The figure shows that dentals, form the biggest class (43%). Looking back at Figure 4.5 we see that dentals are among the most frequent coda consonants in all manner subclasses. Labials are the second biggest group, but their frequency is less than half that of dentals. The percentages of the rest range from 14% (palatals) to 5% (velars).

Figure 4.7: *Frequencies and percentages of the place classes in the coda of the CVC syllables of the DBS*



4.2.1.3. Nucleus

For each vowel, the theoretical maximal frequency in CVC syllables is 529 (=23x23). Figure 4.8 shows the attested frequencies for each vowel. All vowels can occur in the nucleus position, but not in all C_C combinations. We see that the most sonorous vowels /ä/ and /a/ have the highest frequencies, and the low sonorous /u/ has the lowest frequency, with the frequencies of /i/ and the mid vowels in between. However, the frequencies of different vowels do not differ very much here: the percentages range from 20% (for /ä/) to 13% (for /u/). As to the gaps, /ä/ occurs in 66% of the theoretically possible CVC syllables, and even the least frequent vowel /u/ occurs in nearly half (47%) of the theoretical possibilities.

Since individual vowels differ rather little in their frequencies, differences between vowel classes are also quite small. This is seen in Figure 4.9. It shows that the frequency decreases when the sonority decreases, but not steeply. Front vowels are slightly more frequent than back vowels, and there is only a small preference for short vowels in CVC syllables.

Figure 4.8: *Frequencies and percentages of in the nucleus vowels of the CVC syllables of the DBS*

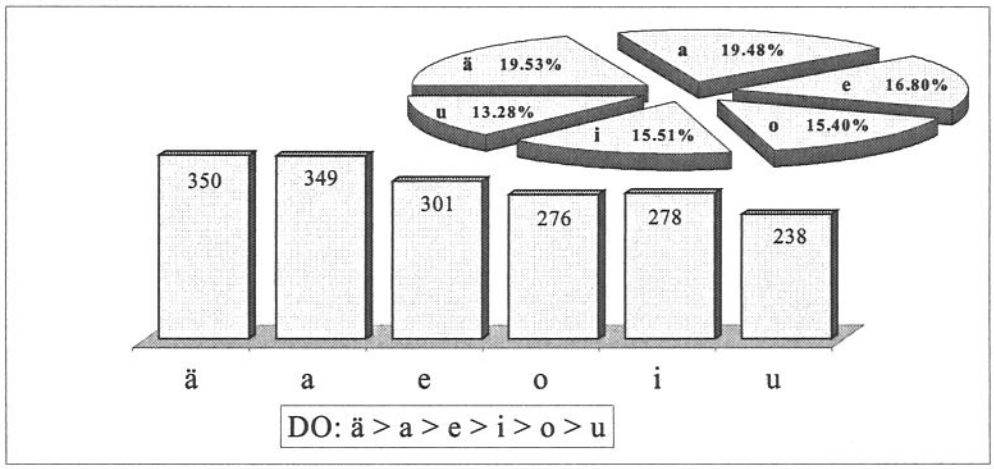
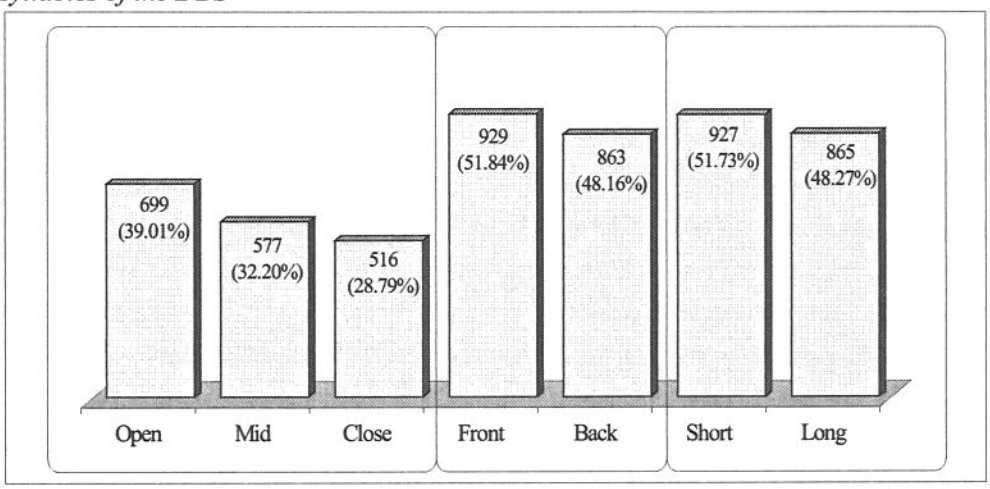


Figure 4.9: *Frequencies and percentage of vowel classes in the nucleus of the CVC syllables of the DBS*



4.2.1.4. Comparison of onset and coda

Figure 4.10 presents consonant frequencies in the onset and the coda. The shaded column indicates that the consonant has a higher frequency in the onset, and the white column means that the frequency is higher in the coda. The figure shows

that stops and affricates (i.e. non-continuants) are usually more frequent in the onset; the only exceptions are the voiced stops /d/ and /G/. Fricatives and sonorants, i.e. continuants, tend to be more frequent in the coda. The only clear exception is /v/, which disfavours the syllable-final position.

Differences between manner subclasses can be seen more clearly in Figure 4.11. It shows that non-continuants are more frequent in the onset. This difference is particularly clear in the affricates. Continuants, except for the glide, are more frequent in the coda. The glide does not show any clear preference.

Figure 4.10: *Comparison of the frequencies of consonants in the onset and coda of CVC syllables of the DBS*

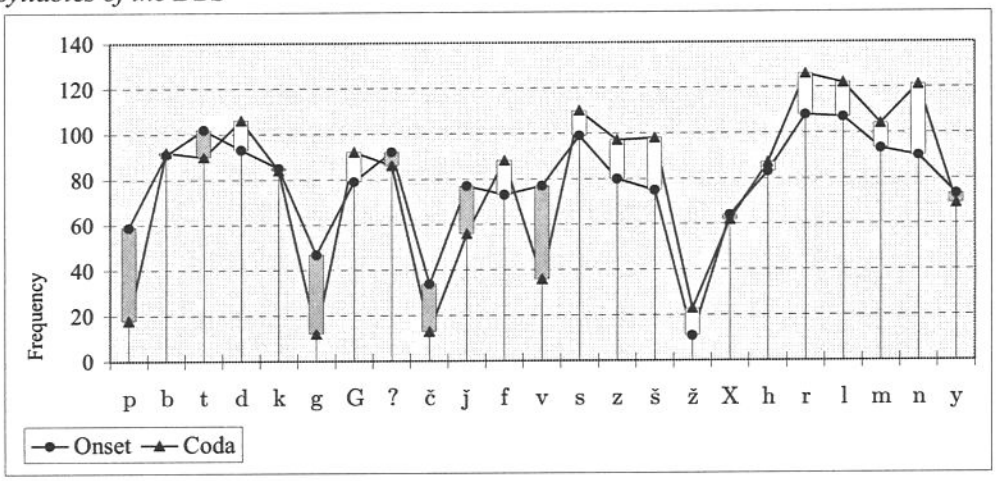


Figure 4.11: *Comparison of the frequencies of the manner subclasses in the onset and coda of the CVC syllables of the DBS*

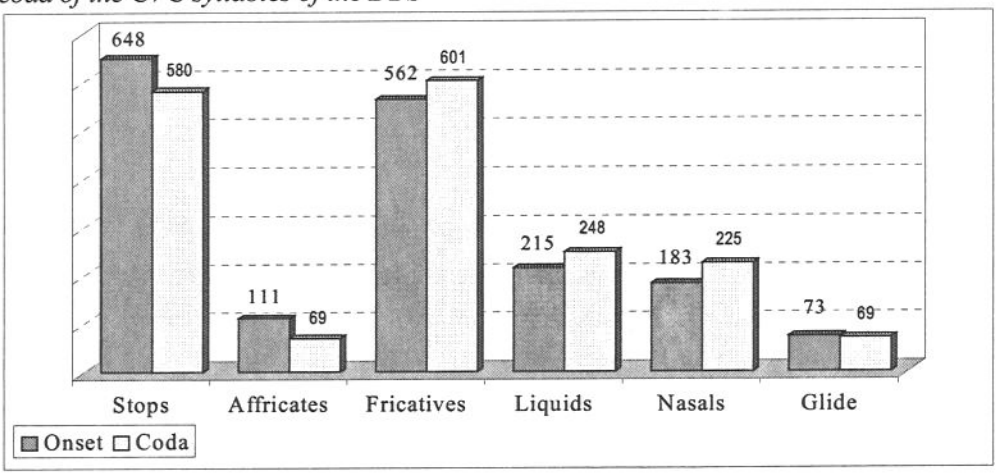


Table 4.1 indicates that the difference between continuants and non-continuants is systematic, also on the level of different nuclear vowels. The table presents onset and coda frequencies of manner subclasses for each nucleus vowel separately. We see that with two exceptions, non-continuants and continuants differ systematically, regardless of the nucleus vowel. The exceptions are fricatives with /i/ in the nucleus, and the glide with most vowels. Whether this tendency of non-continuants to prefer the onset and for continuants to prefer the coda is more general in languages or just an idiosyncratic feature of some of them is not clear. In any case, this tendency does not follow from the sonority hierarchy.

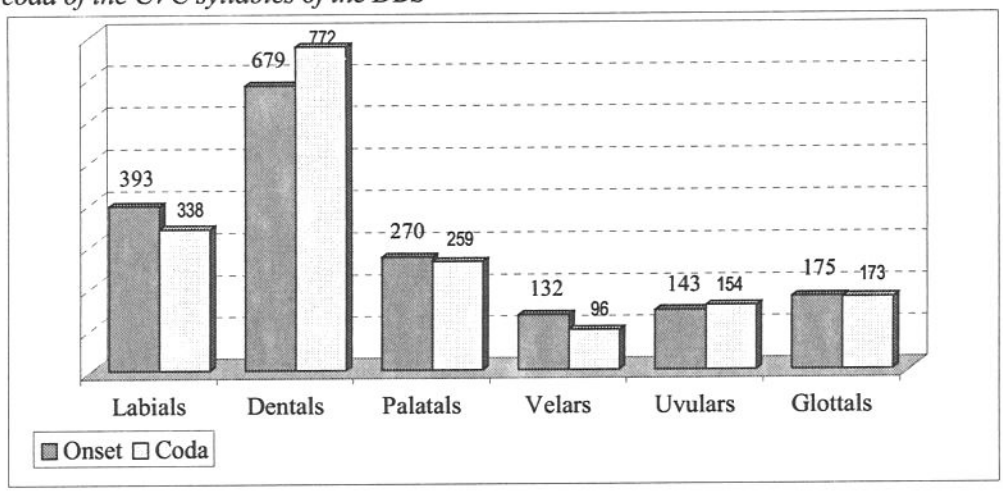
Table 4.1: *Onset and coda consonant frequencies of the manner subclasses for each of the six vowels in CVC syllables of the DBS. (The arrow points to the higher frequency; the sign " ‖ " indicates that the frequency is the same in the onset and the coda.)*

		Non-continuants			Continuants				
Nuc.		**Stops**	**Affricates**	**Total**	**Fricatives**	**Liquids**	**Nasals**	**Glide**	**Total**
ä	**Onset**	129 ↑	23 ↑	152 ↑	111 ↓	37 ↓	37 ↓	13 ‖	198 ↓
	Coda	117 ↓	14 ↓	131 ↓	125 ↑	40 ↑	41 ↑	13 ‖	219 ↑
a	**Onset**	122 ↑	24 ↑	146 ↑	113 ↓	40 ↓	32 ↓	18 ↑	203 ↓
	Coda	113 ↓	17 ↓	130 ↓	119 ↑	45 ↑	43 ↑	12 ↓	219 ↑
e	**Onset**	102 ↑	19 ↑	121 ↑	96 ↓	35 ↓	31 ↓	18 ‖	180 ↓
	Coda	93 ↓	10 ↓	103 ↓	102 ↑	44 ↑	34 ↑	18 ‖	198 ↑
o	**Onset**	105 ↑	13 ↑	118 ↑	80 ↓	35 ↓	36 ↓	7 ↑	158 ↓
	Coda	91 ↓	9 ↓	100 ↓	94 ↑	43 ↑	38 ↑	1 ↓	176 ↑
i	**Onset**	93 ↑	17 ↑	110 ↑	97 ↑	36 ↓	27 ↓	8 ↓	168 ↓
	Coda	88 ↓	11 ↓	99 ↓	84 ↓	40 ↑	38 ↑	17 ↑	179 ↑
u	**Onset**	96 ↑	15 ↑	111 ↑	65 ↓	33 ↓	20 ↓	9 ↑	127 ↓
	Coda	78 ↓	8 ↓	86 ↓	77 ↑	36 ↑	31 ↑	8 ↓	152 ↑

Figure 4.11 and Table 4.1 show that different manner subclasses behave somewhat differently in the onset and coda. There are differences between place classes too, as Figure 4.12 shows, but these differences can be mainly explained on the basis of manner subclasses. We see, for instance, that dentals prefer the coda. Part of the explanation is that five out of the seven dentals are continuants, which generally prefer the coda position. However, we have to remember here

that all dentals, even non-continuants, had relatively high frequencies in the coda. Labials, on the other hand, are more frequent in the onset. Two of the five labials are stops, which generally prefer the onset, and one labial continuant, /v/, exceptionally prefers the onset. We also see that velars are more frequent in the onset. This is because all velars are stops, which prefer the onset. Palatals, uvulars, and glottals do not have a clear preference. This is due to the fact that each group has an equal number of continuants and non-continuants, whose preferences are the opposite.

Figure 4.12: *Comparison of the frequencies of the place classes in the onset and the coda of the CVC syllables of the DBS*



4.2.2. CVC syllables from the etymological point of view

Out of the 1792 CVC syllables in the Data of Basic Syllables, 916 (=51%) come from words of Persian origin, and the remaining 876 (49%) syllables originate in loanwords. The biggest contributor is Arabic with 680 syllables. A small number of syllables, 88 altogether, are multi-contributor.

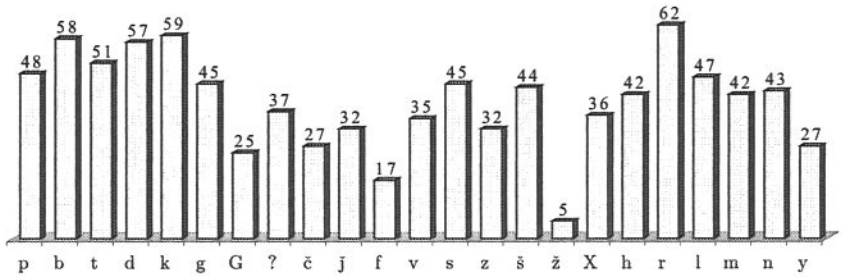
In the following, I shall characterise the properties of CVC syllables in words of Persian origin (DPS) and make comparisons with CVC syllables of foreign words (DNPS). I shall deal with each syllabic position separately.

4.2.2.1. Onset

DPS

Figure 4.13 presents consonant frequencies in the onset of the CVC syllables in the DPS. It shows that all consonants can occur in the onset, but none of them in all _VC combinations. Since the theoretically maximal frequency for each CVC onset consonant in DBS is 138, and since DPS constitutes 51% of DBS, we could expect that 70 (= 51% of 138) is the theoretically maximal average frequency of CVC onset consonants in DPS. As Figure 4.13 shows, all onset consonant frequencies are below this average maximum. Closest to the maximal frequency are the stops /b/, /d/ and /k/, and the liquid /r/, with a frequency more than 80% of the maximum. The lowest frequency belongs to /ž/, with a frequency less than 10% of the maximum.

Figure 4.13: *Frequencies and percentages of consonants in the onset of the CVC syllables of the DPS*

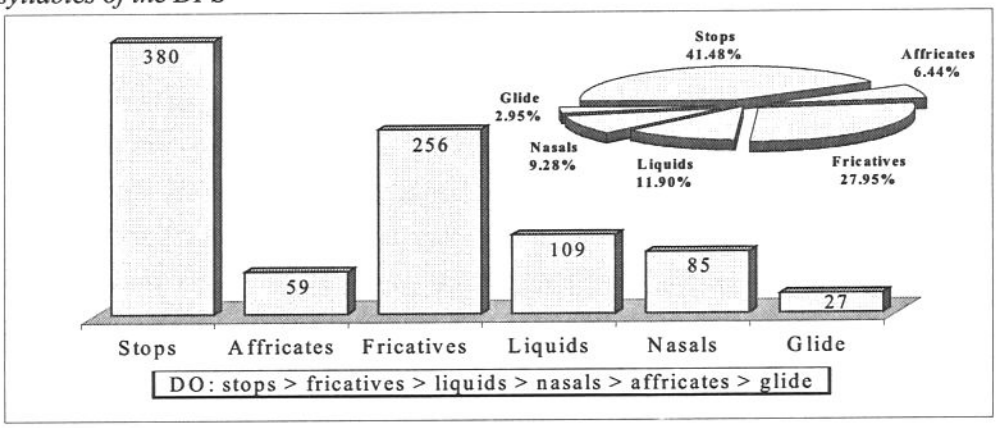


p	5.24%	d	6.22%	G	2.73%	j	3.49%	s	4.91%	ž	0.55%	r	6.77%	n	4.69%
b	6.33%	k	6.44%	?	4.04%	f	1.86%	z	3.49%	χ	3.93%	l	5.13%	y	2.95%
t	5.57%	g	4.91%	č	2.95%	v	3.82%	š	4.80%	h	4.59%	m	4.59%	100.00%	

DO: r > k > b > d > t > p > l > g, s > š > n > h, m > ? > χ > v > j, z > č, y > G > f > ž

Figure 4.14 presents onset consonant frequencies and percentages of different manner subclasses in the onset position of CVC syllables in the DPS. The big proportion of stops (41%) is conspicuous. Together with affricates, they cover nearly half (48%) of all onset consonant tokens in the DPS. Fricatives have a proportion of 28%. The percentages of the rest range between 12% (liquids) and 3% (the glide).

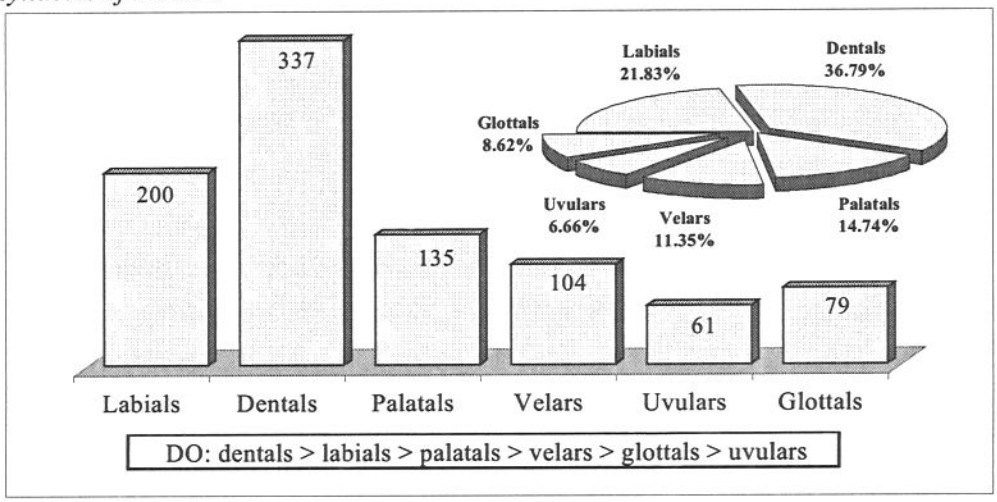
Figure 4.14: *Frequencies and percentages of manner subclasses in the onset of the CVC syllables of the DPS*



Frequencies and percentages of different place classes in the onset of the DPS can be seen in Figure 4.15. Dentals, with a proportion of 37%, are the biggest group, and labials (22%) are the next biggest group. Velars have a clearly bigger percentage in the DPS (11%) than in the DBS (7%).

Obstruents form 75.87% and sonorants 24.13% of the consonants in the onset. 58.42% of the obstruents are voiceless and 41.58% voiced.

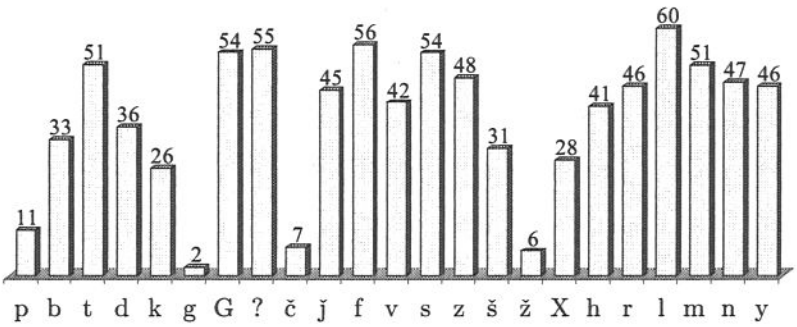
Figure 4.15: *Frequencies and percentages of place classes in the onset of the CVC syllables of the DPS*



DNP: comparison of the DPS and DNPS

Figure 4.16 shows onset consonant frequencies in the CVC syllables of non-Persian origin. We see that here, too, all consonants can occur in the onset, but with varying frequencies. Since the DNPS covers 49% of the DBS, the theoretically maximal (average) frequency for each consonant is 68. Among those that come closest to the maximal frequency are the stops /G/ and /ʔ/, the fricatives /f/ and /s/, and the liquid /l/, with roughly 80% or more of the maximum. On the other hand, there are three consonants whose frequency is below 10% of the maximum, namely, /g/, /č/, and /ž/.

Figure 4.16: *Frequencies and percentages of consonants in the onset of the CVC syllables of the DNPS*



p	1.26%	d	4.11%	G	6.16%	j	5.14%	s	6.16%	ž	0.68%	r	5.25%	n	5.37%
b	3.77%	k	2.97%	?	6.28%	f	6.39%	z	5.48%	χ	3.20%	l	6.85%	y	5.25%
t	5.82%	g	0.23%	č	0.80%	v	4.79%	š	3.54%	h	4.68%	m	5.82%	100.00%	

DO: l>f>ʔ>G,s>t,m>z>n>r,y>j>v>h>d>b>š>χ>k>p>č>ž>g

As Figures 4.13 and 4.16 indicate, the onset consonant frequencies of CVC syllables differ in the DPS and DNPS. We can see these differences more clearly in Figure 4.17 which compares the frequencies in the two data. The white column indicates that the consonant frequency is higher in the DPS, and the grey column indicates the opposite, i.e. that the frequency is higher in the DNPS.

Figure 4.17: *Comparison of consonant frequencies in the onset of the CVC syllables of the DPS and DNPS*

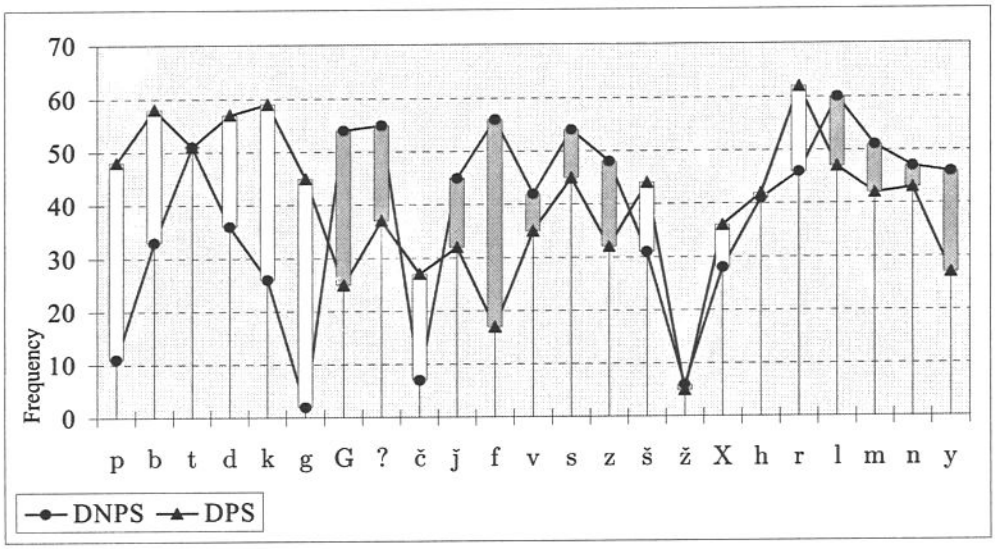


Figure shows that stops tend to have higher frequencies in the DPS than in the DNPS. The difference is conspicuous in /p/ and /g/. However, there are exceptions: /t/ has an equal frequency in both, and /G/ and /ʔ/ have higher frequencies in the DNPS. Fricatives have higher frequencies in the onset than in the DNPS, except for /š/ and /χ/, which are more frequent in the DPS. Sonorants have higher frequencies in the DNPS, except for /r/, which is more frequent in the DPS.

Figure 4.18 shows the percentages of onset consonant tokens in CVC syllables of Persian and foreign origin. As can be expected on the basis of Figure 4.17, stop consonant tokens, except for /G/ and /ʔ/, tend to come more from etymologically Persian syllables. Thus, 96% of the occurrences of /g/ and 81% of the tokens of /p/ come from syllables of Persian origin, while only 32% of the tokens of /G/ and 40% of the occurrences of /ʔ/ come from etymologically Persian syllables. The Persian share is also big in /č/, almost 80%. The foreign share is biggest in /f/, almost 77%.

Figure 4.18: *Percentages of onset consonant tokens coming from Persian vs. foreign sources in the CVC syllables of the DBS*

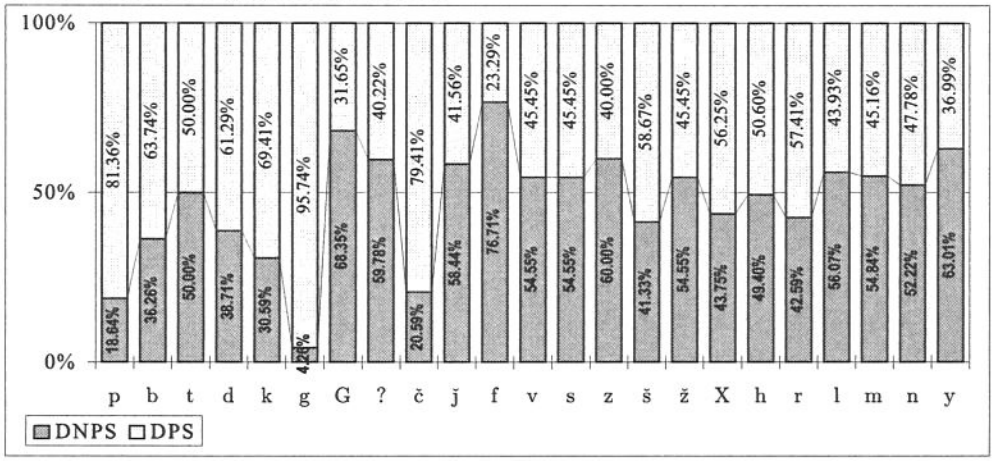


Figure 4.19 shows what percentages of manner subclasses in the onset of CVC syllables in the DBS come from etymologically Persian syllables and what percentages from etymologically non-Persian syllables. We see that in Persian, stops have the biggest proportion, (59%), and the glide has smallest proportion (37%). In the other subclasses, the etymologically Persian share varies between 46% and 53%.

Figure 4.19: *Percentages of different manner subclasses coming from Persian vs. foreign sources in the onset of CVC syllables of the DBS*

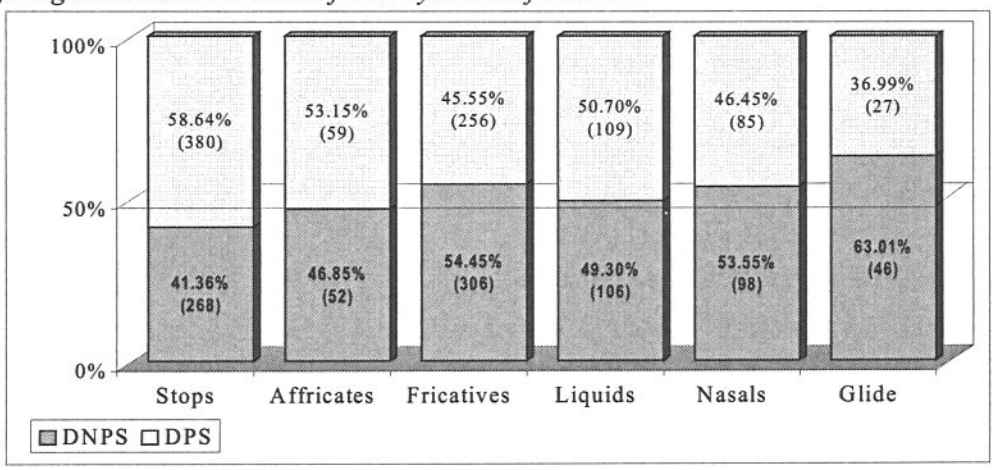
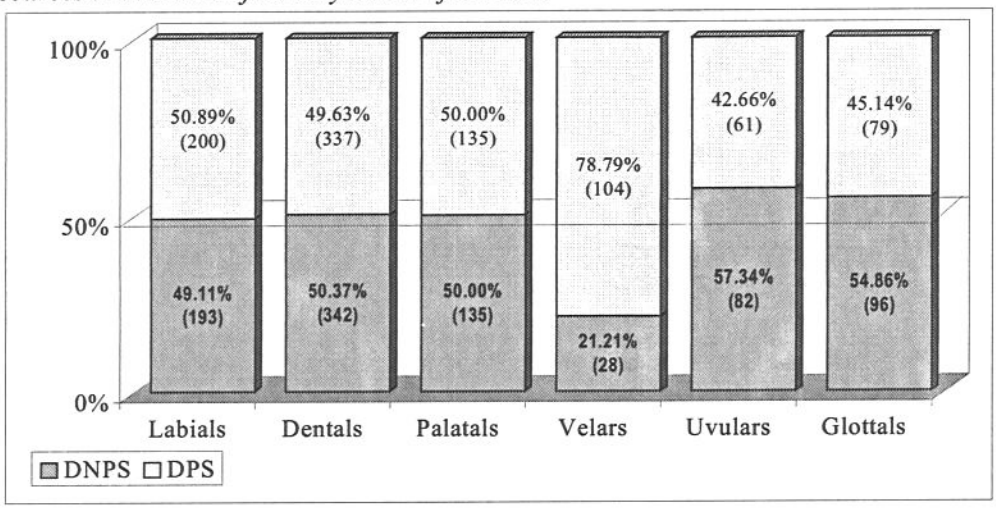


Figure 4.20 indicates what percentages of place classes in the onset of CVC syllables in the DBS come from etymologically Persian syllables and what percentages from the etymologically non-Persian syllables. The figure shows that velars deviate most from the expected 51% of Persian origin to 49% of foreign origin, 79% of velar occurrences come from words of Persian origin. On the other hand, the foreign proportion is clearly above the average 49% in uvulars and glottals.

Figure 4.20: *Percentages of different place classes coming from Persian vs. foreign sources in the onset of CVC syllables of the DBS*



4.2.2.2. Coda

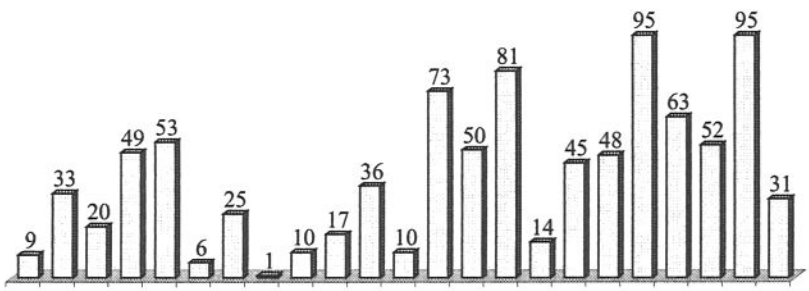
DPS

Coda consonant frequencies and percentages in the DPS are given in Figure 4.21. The figure shows that all consonants occur syllable finally in the DPS, but the frequencies are quite unevenly distributed in the coda, ranging between one for /ʔ/[42] and 95 for /r/ and /n/. If we think that here, as in the onset, the average

[42] The consonant /ʔ/ occurs only in the word /bäʔ-bäʔ/, which is an onomatopoeic word meaning 'baaing, bleating'. With the exception of this example, all other occurrences of /ʔ/ in the coda position come from etymologically Arabic words. In Modern Persian, this

theoretical maximal frequency is 70, we see that four consonants exceed this frequency, namely, /s/, /š/, /r/ and /n/. On the other hand, several consonants have low frequencies, at most 20% of the theoretical (average) maximum, e.g. /ʔ/, /g/, /p/, /č/, /v/, and /ž/.

Figure 4.21: *Frequencies and percentages of the consonants in the coda of the CVC syllables of the DPS*



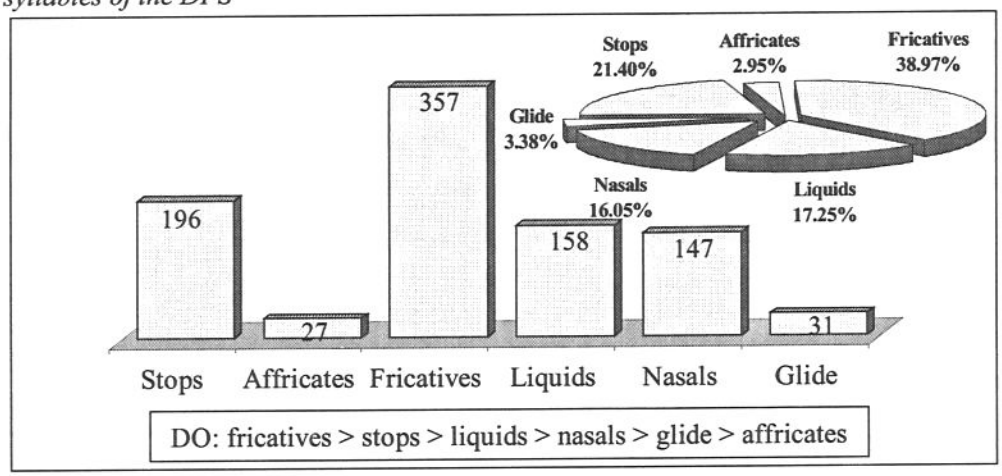
p	0.98%	d	5.35%	G	2.73%	j	1.86%	s	7.97%	ž	1.53%	r	10.37%	n	10.37%
b	3.60%	k	5.79%	ʔ	0.11%	f	3.93%	z	5.46%	χ	4.91%	l	6.88%	y	3.38%
t	2.18%	g	0.66%	č	1.09%	v	1.09%	š	8.84%	h	5.24%	m	5.68%	100.00%	

DO: r>n>š>s>l>k>m>z>d>h>χ>f>b>y>G>t>j > ž >č > v>p>g>ʔ

Of the manner subclasses, stops and affricates, i.e. non-continuant consonants, tend to have low frequencies, while continuants tend to be frequent in the coda. This can be seen clearly in Figure 4.22, which presents the frequencies and percentages in different manner subclasses. The proportion of stops and affricates together is about one-fourth only, while the remaining three-fourths are taken by continuants.

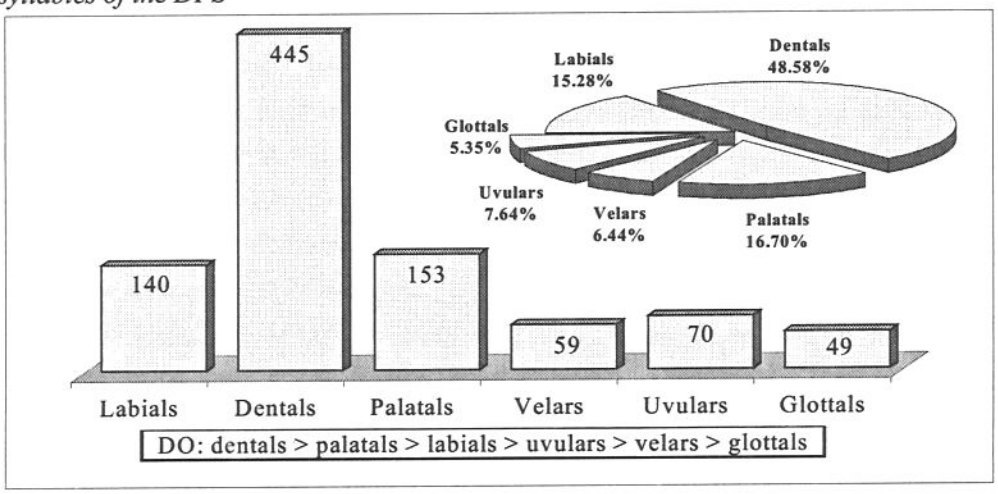
word is undergoing a steady process where the glottal stop /ʔ/ is substituted with a prolonged /ä/.

Figure 4.22: *Frequencies and percentages of manner subclasses in the coda of the CVC syllables of the DPS*



The frequencies and percentages of coda consonants in different place classes are given in Figure 4.23. The big proportion of dentals, nearly half of all consonant tokens, is conspicuous. But here, as in the onset, the explanation can be found in the manner group since the class of dentals includes several continuants, which tend to have high frequencies in the coda.

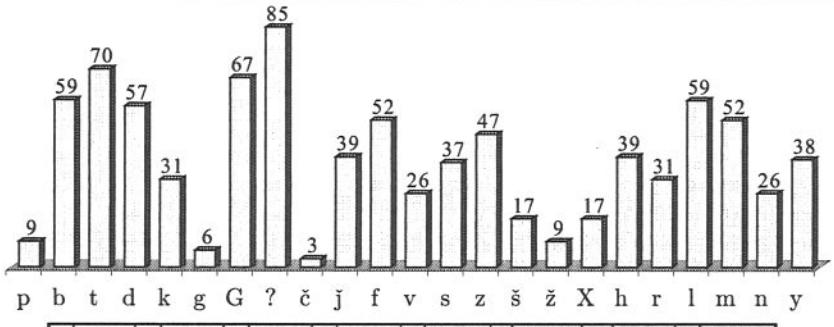
Figure 4.23: *Frequencies and percentages of place classes in the coda of the CVC syllables of the DPS*



DNPS: comparison of the DPS and DNPS

Figure 4.24 shows coda consonant frequencies and percentages in the etymologically non-Persian data. Here, as in the DPS, frequencies are unevenly distributed, but in a different way: several stops have high frequencies, while sonorants tend to be less frequent. Thus, those that almost reach or exceed the (average) theoretical maximum of 68 are all stops: /t/, /G/, and /ʔ/.

Figure 4.24: *Frequencies and percentages of consonants in the coda of the CVC syllables of the DNPS*



p	1.03%	d	6.51%	G	7.65%	j	4.45%	s	4.22%	ž	1.03%	r	3.54%	n	2.97%
b	6.74%	k	3.54%	ʔ	9.70%	f	5.94%	z	5.37%	χ	1.94%	l	6.74%	y	4.34%
t	7.99%	g	0.68%	č	0.34%	v	2.97%	š	1.94%	h	4.45%	m	5.94%	100.00%	

DO: ʔ>t>G>b,l>d>f,m>z>j,h>y>s>k,r>v,n>š>χ>p,ž>g>č

Differences in coda frequencies between the DPS and the DNPS can be seen in Figure 4.25. The shaded column means that the frequency is higher in the DNPS, while a white column indicates that the DPS has a higher frequency. As the figure shows, non-continuants usually have a higher frequency in the coda of syllables of non-Persian origin; the differences are particularly big in /t/, /G/ and /ʔ/. The only exceptions are /k/ and /č/, which have higher frequencies in the DPS. Continuants, on the other hand, tend to have higher frequencies in the data of Persian origin; the differences are conspicuous in /š/, /r/ and /n/ in particular. Exceptions to this tendency are the labial fricatives and the glide.

Figure 4.25: *Comparison of the frequencies of coda consonants in the CVC syllables of the DPS and DNPS*

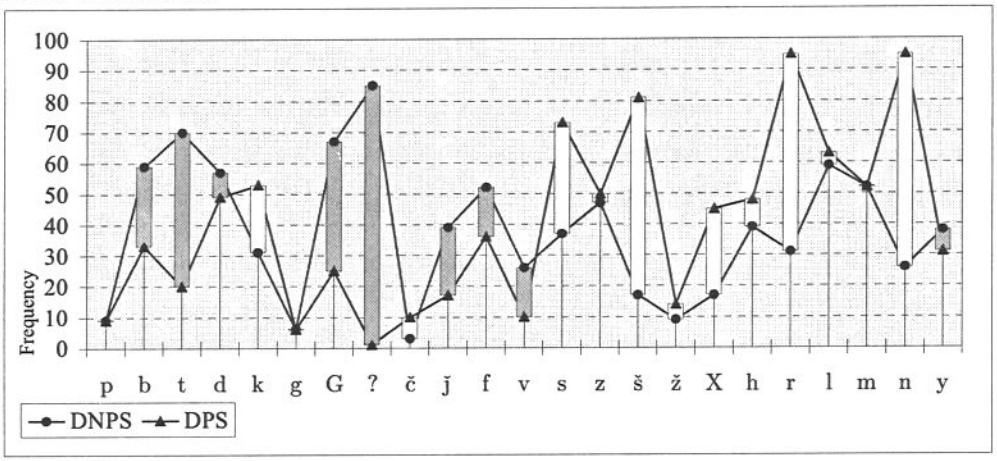


Figure 4.26 presents percentages of coda consonant tokens coming from etymologically Persian syllables vs. non-Persian syllables. The figure shows that the foreign proportion is substantial in /ʔ/ with practically all tokens coming from non-native syllables. The foreign proportion is more than 70% in /t/, /G/ and /v/. On the other hand, over 70 % of the tokens of the following consonants come from syllables of Persian origin: /č/, /š/, /r/ and /n/.

Figure 4.26: *Percentages of coda consonant tokens coming from Persian vs. non-Persian sources in the CVC syllables of the DBS*

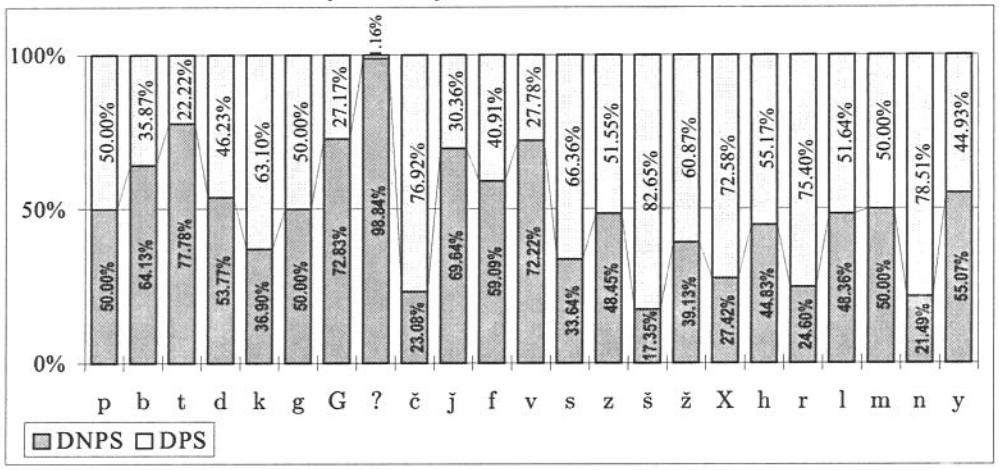


Figure 4.27 shows the percentages of coda consonants coming from etymologically Persian vs. non-Persian data in different manner subclasses. The figure indicates what we could expect already (on the basis of Figure 4.25, for example): the native proportion is bigger in continuants (liquids, nasals and fricatives), while the foreign share is bigger in non-continuants (stops and affricates), but also in the glide.

Figure 4.27: *Percentages of different manner subclasses coming from Persian vs. foreign sources in the coda of CVC syllables of the DBS*

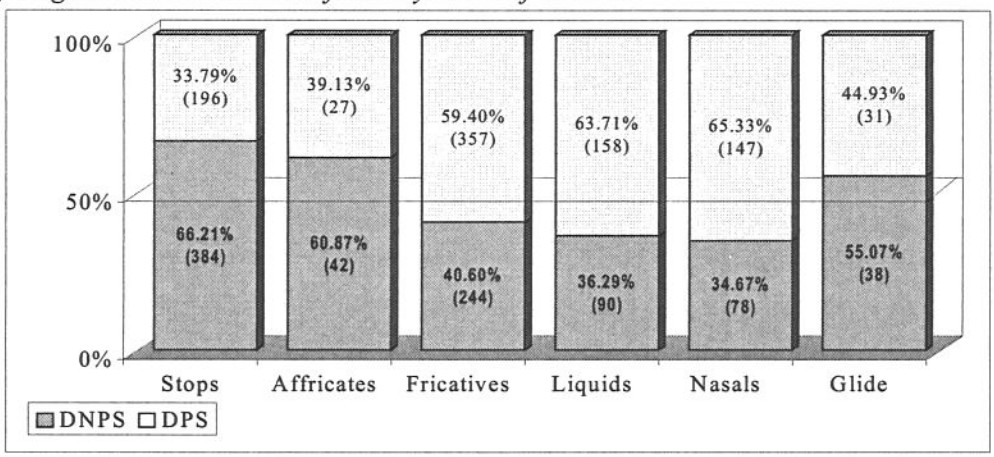
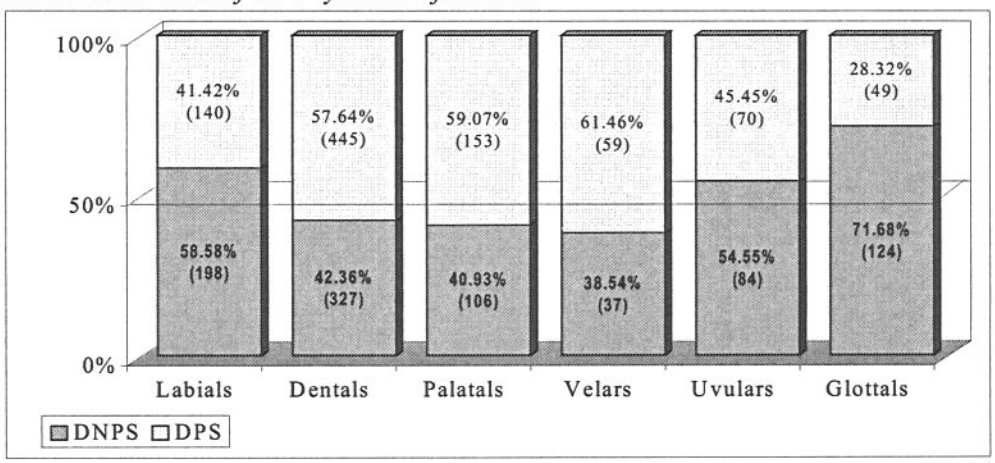


Figure 4.28: *Percentages of different place classes coming from Persian vs. foreign sources in the coda of CVC syllables of the DBS*



The Persian vs. non-Persian proportions of coda consonant tokens in different place classes are given in Figure 4.28. It shows that all place classes

deviate from the average 51% Persian to 49% non-Persian. In dentals, palatals and velars, the Persian shares range from 58% to 61%. In the rest of the place classes, the proportion of Persian is less than 50%; of labials and uvulars, the Persian share is somewhat over 40%, while only 28% of coda glottals come from words of Persian origin.

4.2.2.3. Nucleus

DPS

As mentioned before, the DPS contains 916 syllables, and therefore also 916 vowel tokens. Vowel frequencies are presented in Figure 4.29. Open vowels differ clearly from the rest in that they have the highest frequencies, and it is the back vowel, /a/, that has the biggest proportion, 25%. What is also striking is that mid vowels have the lowest frequencies. However, their frequencies are not much lower than those of close vowels. If all the vowels are treated as one group, there is a correlation between vowel frequency and sonority, but not a high one. However, the long vowels /a, i, u/ and short vowels /ä, e, o/ seem to function as two separate groups. In each group, vowel frequency decreases when sonority decreases.

Figure 4.29: *Frequencies and percentages of vowels in the nucleus of the CVC syllables of the DPS*

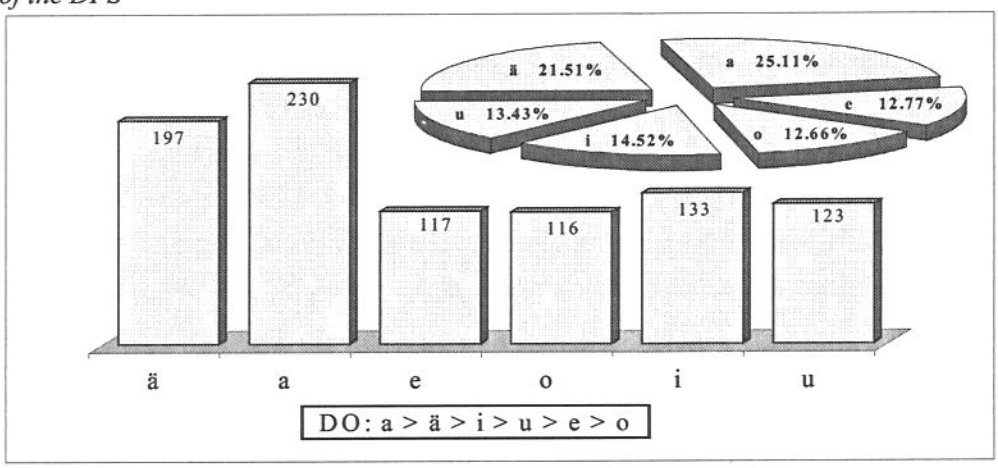
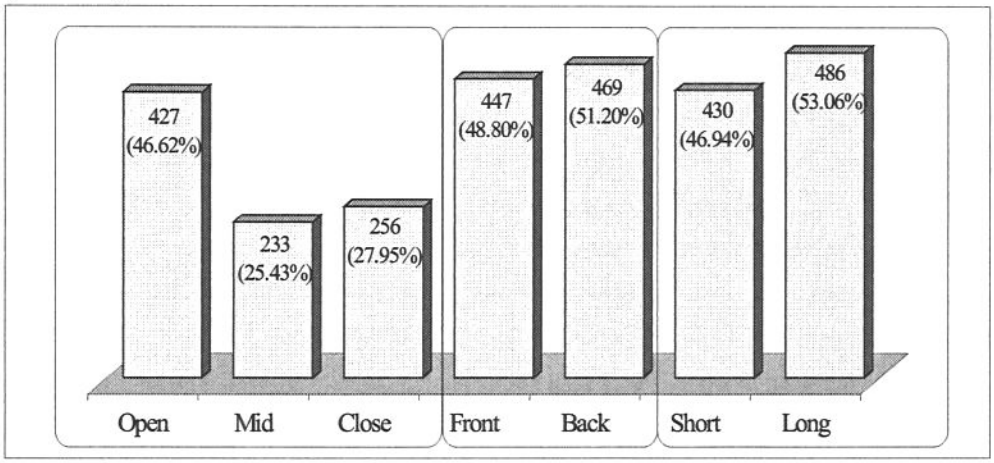


Figure 4.30 gives frequencies and percentages of different vowel classes in the CVC syllables of the DPS. The figure shows that open vowels comprise nearly half (47%) of all vowel occurrences, and close vowels (28%) have a somewhat bigger percentage than mid vowels (25%). Moreover, back vowels are slightly more frequent than front vowels and long vowels have a bigger percentage than short vowels.

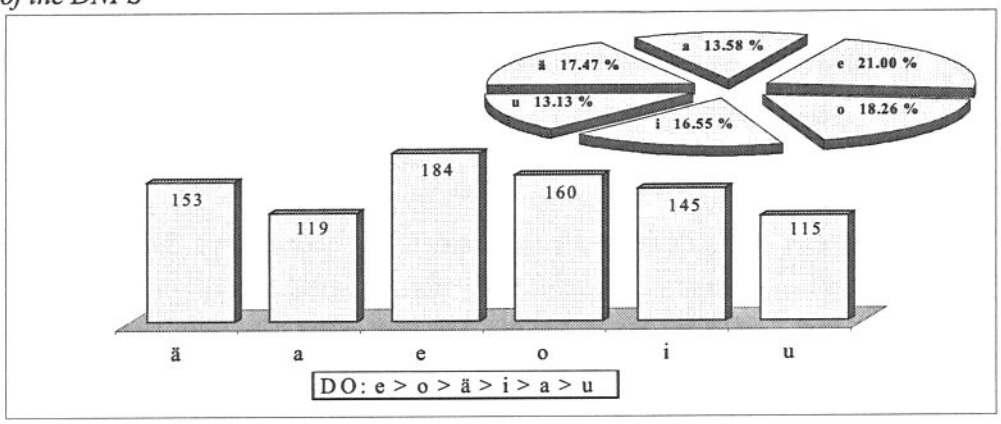
Figure 4.30: *Frequencies and percentages of vowel classes in the nucleus of the CVC syllables of the DPS*



DNPS: Comparison of the DPS and DNPS

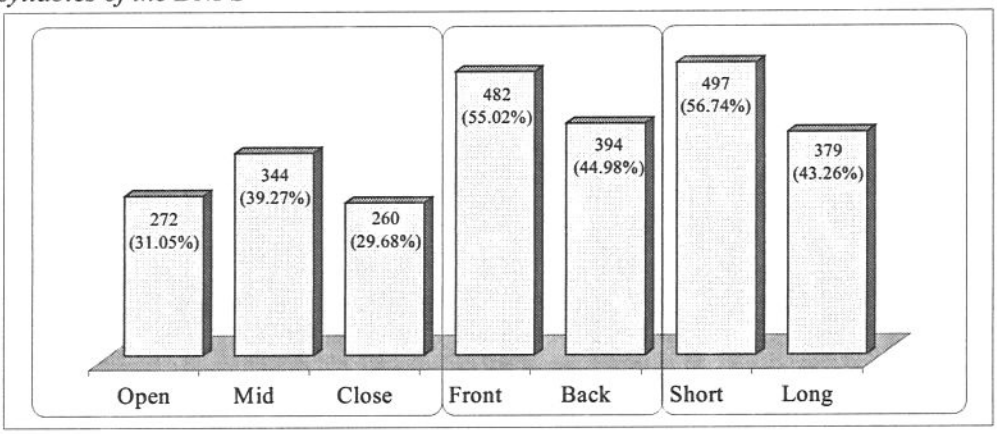
Figure 4.31 presents frequencies and percentages of vowels in the basic CVC syllables coming from foreign sources. The figure shows that vowel frequencies are somewhat more evenly distributed here than in the DPS (see Figure 4.29), i.e. the percentages of vowels in Figure 4.31 vary from 21% for /e/ to 13% for /u/. Moreover, the highest frequencies belong to the mid vowels, and especially to /e/, while the lowest frequencies belong to one open vowel, /a/, and one close vowel, /u/.

Figure 4.31: *Frequencies and percentages of vowels in the nucleus of the CVC syllables of the DNPS*



The proportions of different vowel classes can be seen better in Figure 4.32. It shows that the classes of open and close vowels each have a proportion of roughly 30%, while mid vowels have a proportion of nearly 40%. This means that only mid and close vowels show a positive correlation between vowel frequency and sonority; the small proportion of open vowels violates the correlation. Looking back at Figure 4.31, we see that if short vowels (/ä, e, o/) and long vowels (/a, i, u/) are treated separately, the correlation does not hold for them either. The figure also shows that front vowels are somewhat more frequent than back vowels, and the percentage of short vowels (57%) is bigger than the percentage of long vowels (45%).

Figure 4.32: *Frequencies and percentages of vowel classes in the nucleus of the CVC syllables of the DNPS*



When comparing the vowel profiles of CVC syllables in the DPS (Figures 4.29 and 4.30) and in the DNPS (Figures 4.31 and 4.32), we see that the results of the two data are to some extent opposite. To mention just a few examples: /a/ has a higher frequency than /ä/ in the DPS, but in the DNPS it is the other way round; in the DPS, the class of mid vowels has the lowest frequency, but it has the highest frequency in the DNPS; in the DPS, long vowels are more frequent than short vowels, but in the DNPS the situation is the opposite. Thus, when the DPS and the DNPS are united to form the DBS, frequency differences between vowels and vowel classes tend to diminish. The result is what was seen before in the DBS (see Figures 4.8 and 4.9), where the frequencies of vowels and vowel classes differ relatively little.

Finally, let us see what percentages of the tokens of each vowel in the CVC syllables of the DBS come from Persian vs. foreign sources. This is presented in Figure 4.33. It shows that /u/ comes closest to the ratio 51% of Persian origin to 49% of foreign origin. The Persian share is particularly big in the open vowels (66% for /a/; 56% for /ä/). On the other hand, mid vowels in particular have a relatively big foreign share (61% for /e/; 58% for /o/). The percentages of vowel classes from Persian vs. foreign sources are given in Figure 4.34. As can be expected, the Persian share is more than the average 51% in the class of open vowels, but also in long and back vowels. On the other hand, the Persian share is well below the average in mid vowels.

Figure 4.33: *Percentages of Persian vs. non-Persian sources of nuclear vowels in CVC syllables of the DBS*

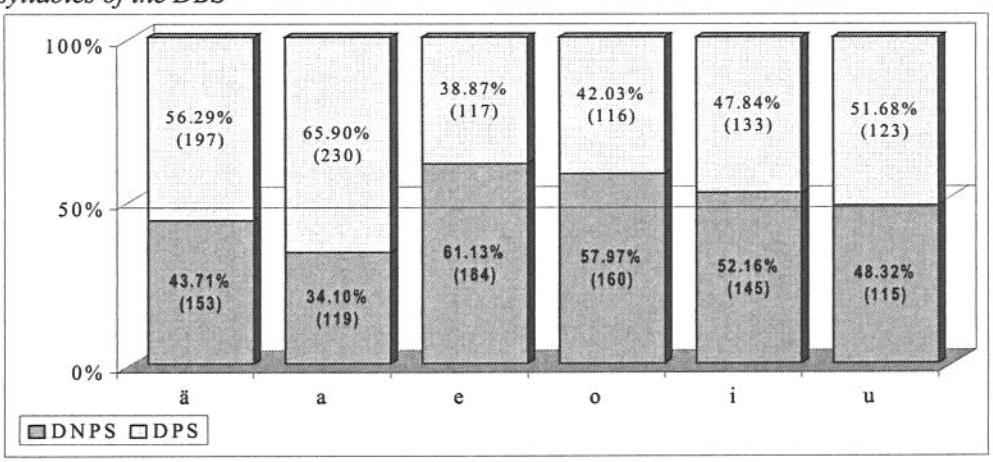
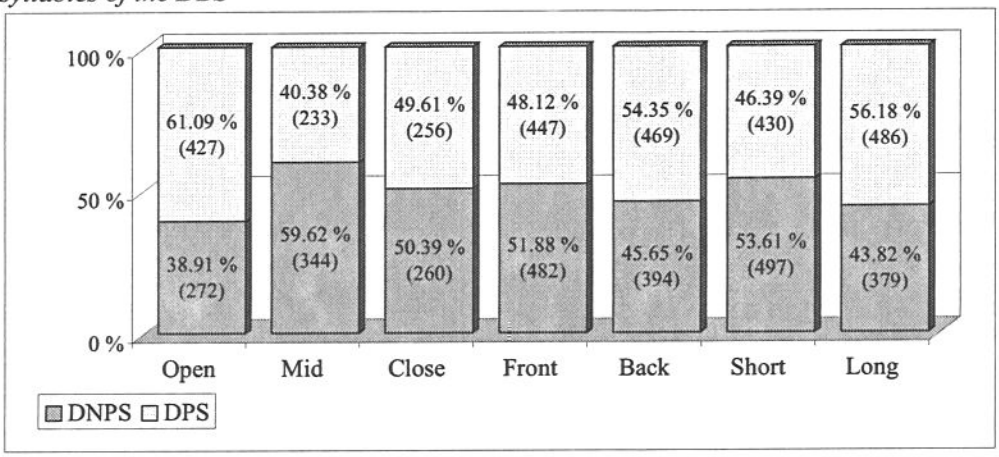


Figure 4.34: *Percentages of Persian vs. non-Persian sources in vowel classes in CVC syllables of the DBS*



4.2.2.4. Comparison of onset and coda

DPS

We saw in 4.2.1.4 that non-continuants in the DBS tend to be more frequent in the onset of CVC syllables, and continuants, with the exception of the glide, tend to occur more in the coda. This tendency can be seen in the etymologically Persian data and to an even greater extent. Figure 4.35 compares consonant frequencies in the onset and coda in the DPS. The shaded column means that the frequency is higher is the onset, and the white column indicates that the frequency is higher in the coda. We see that all other stops except /G/, and both affricates have a higher frequency in the onset. /G/ is as frequent in the onset as in the coda. All other continuants except /v/ have a higher frequency in the coda. Here, as in the DBS, /v/ is rare in the syllable-final position. This implies that non-continuants and continuants in the DPS differ clearly in their preferences, as Figure 4.36 shows. The difference is particularly clear in the non-continuants, whose frequency in the onset is roughly twice as high as their frequency in the coda. Here, as in the DBS, the glide does not show any clear positional preference.

Figure 4.35: *Comparison of consonant frequencies in the onset and the coda of CVC syllables of the DPS*

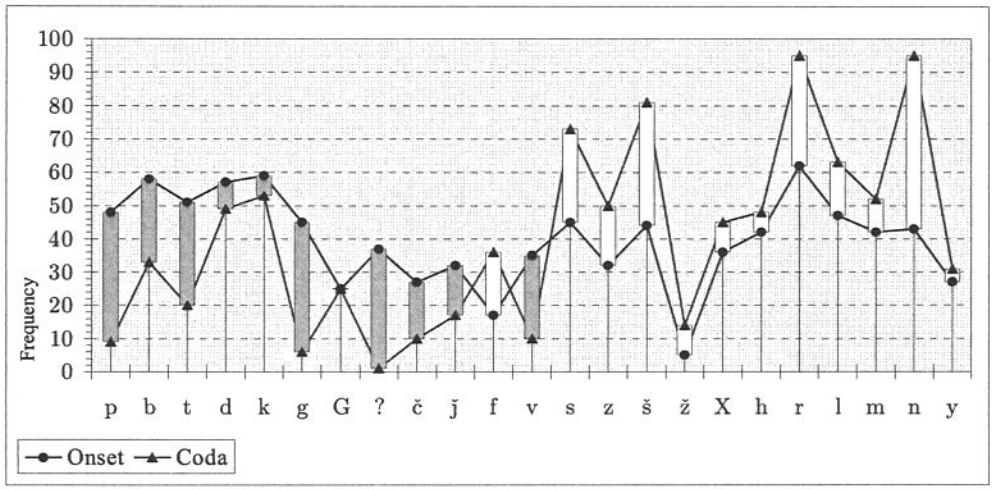
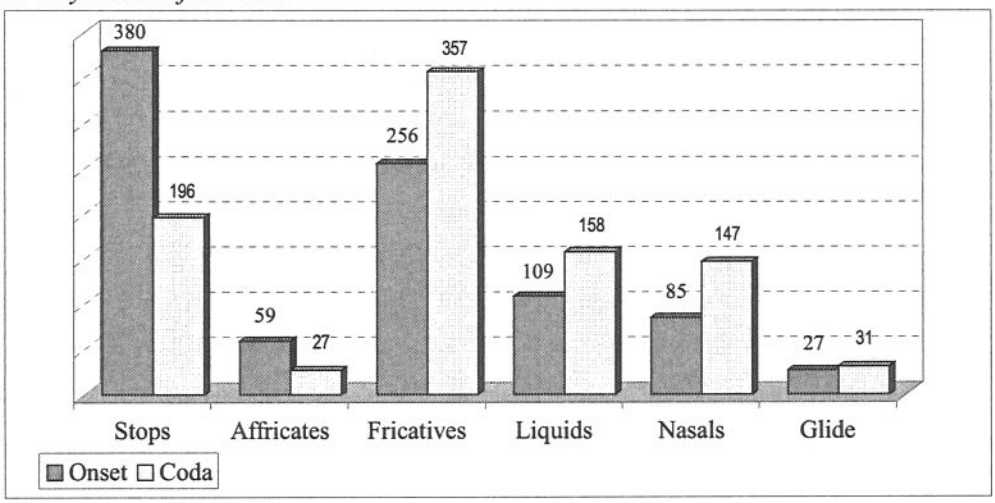


Figure 4.36: *Comparison of manner subclass frequencies in the onset and the coda of CVC syllables of the DPS*



DNPS: Comparison of the DPS and DNPS

Figure 4.37 compares consonant frequencies in the onset and the coda of CVC syllables in the DNPS. A white column means that the consonant is more frequent in the coda, and a shaded column indicates that the consonant is more

frequent in the onset. The figure shows that all stops except /p/ are more frequent in the coda; /p/ shows a minimal preference to the onset. All the rest, except /ž/, are either more frequent in the onset or show no preference at all; /ž/ is slightly more frequent in the coda.

Figure 4.37: *Comparison of consonant frequencies in the onset and the coda of CVC syllables of the DNPS*

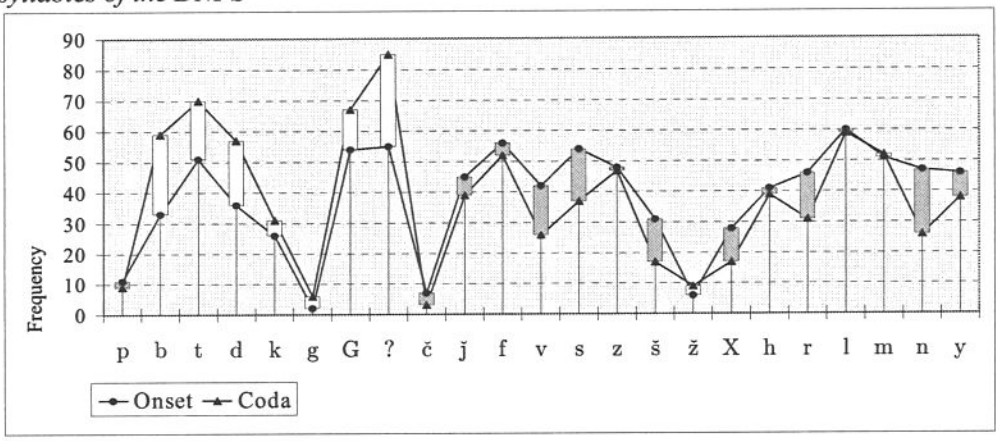
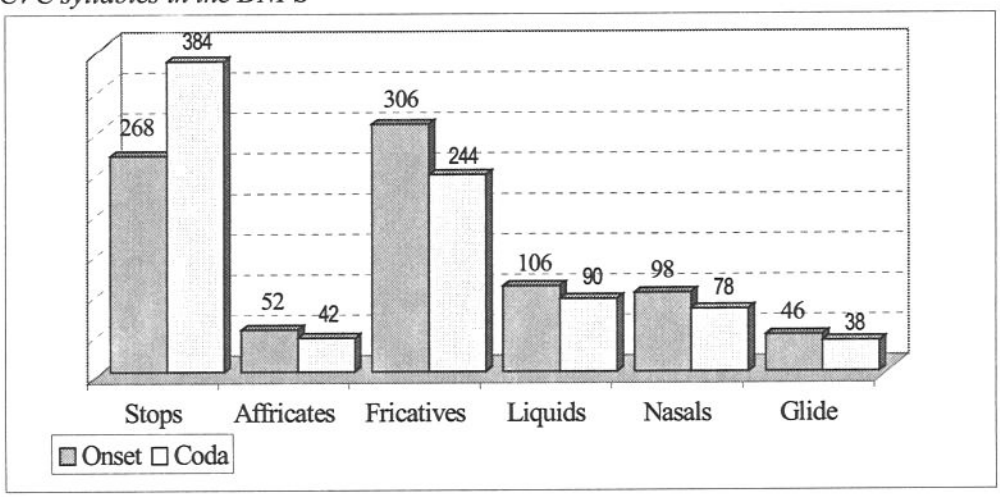


Figure 4.38 gives frequencies of manner subclasses in the onset vs. the coda of CVC syllables in the DNPS. The figure shows that stops clearly prefer the coda, all the rest, i.e. affricates and continuants, prefer the onset.

Figure 4.38: *Comparison of manner subclass frequencies in the onset and the coda of CVC syllables in the DNPS*



A comparison of Figures 4.37 and 4.38, based on the DNPS, to Figures 4.35 and 4.36, based on the DPS, reveals that the two data show opposite tendencies, i.e. in the DPS, non-continuants prefer the onset and continuants prefer the coda, whereas in the DNPS, the biggest non-continuant class, stops, prefers the coda, and the rest prefer the onset. The result of these opposite tendencies is that the DBS, as seen in Figures 4.10 and 4.11, shows the same tendencies as the DPS, but to a much smaller extent.