3. Statistical survey of the Data of Basic Syllables (DBS)

3.1. General features of the DBS

It was mentioned in the previous chapter that the DW contains 24135 syllable tokens. The number of different syllables, however, is as low as 2701. They are called basic syllables, and they form the Data of Basic Syllables (DBS). Thus, each syllable in this data occurs only once.

In terms of etymology, the DBS is divided into two groups: the Data of Persian Syllables (DPS) and the Data of Non-Persian Syllables (DNPS). If a syllable is found in an etymologically Persian word in the DW, it belongs to the DPS even if it is also found in a word of non-Persian origin. Only those syllables that are not found in etymologically Persian words belong to the DNPS.

The following list shows the etymological sources of the basic syllables, the number of syllables that each source has contributed, and the percentage.

Source	**No. of syllables**	**%**
Persian	1315	48.69
Arabic	1091	40.39
French	162	6.00
Multi-contributor	92	3.41
Turkish	25	0.93
English	11	0.41
Russian	2	0.07
Hindi	1	0.03
Greek	1	0.03
Chinese	1	0.03
Total	2701	100.00

Some of the basic syllables are shared by more than one contributing language in the DW. They are called multi-contributor syllables. For example, the syllable /del/ is found in etymologically Persian, Arabic, French, Greek and English words. Most of such syllables are included in the DPS. In the list above, only those multi-contributor syllables that belong to the DNPS are listed as a separate group.

Table 3.1 shows the number of different types of basic syllables (CV, CVC, CVCC) and their percentages in the DBS, as well as in the DPS and the DNPS. The table shows that CVC is the most common type of basic syllable in every group of data: its percentages range from 63% to almost 70%. The next common basic syllable type is CVCC, with percentages ranging from 20% to 36%. Its proportion is lowest in the etymologically Persian data and highest in the etymologically non-Persian data. The number of CVCC syllables in the DNPS is 505, i.e. nearly twice as many as in the DPS, which is 267. The least common basic syllable type is CV: its proportion is 10% of the DPS, 5% of the DBS, and less than a half percent of the DNPS.

Table 3.1: *Frequencies of different syllable types in the DPS, DNPS and DBS*

	DPS		**DNPS**		**DBS**	
	No. of syllables	**Percent**	**No. of syllables**	**Percent**	**No. of syllables**	**Percent**
CV	132	10.04%	5	0.36%	137	5.04%
CVC	916	69.66%	876	63.20%	1792	66.43%
CVCC	267	20.30%	505	36.44%	772	28.53%
Total	1315	100.00%	1386	100.00%	2701	100.00%

Since Persian has 23 consonants and 6 vowels, the maximal number of basic CV syllables that could occur in Modern Persian is 138 (= 23x6). Likewise, the maximal number of basic CVC syllables is 4416 (= 23x6x23), and the theoretical maximum of different CVCC syllables is 73002 (= 23x6x23x23). This makes 77556 possible basic syllables altogether. As we have seen, only a fraction of these, 2701 basic syllables, are found in the data. Table 3.2 compares the actual numbers of basic syllables found in the DBS to the theoretical maximum. The table shows that all but one of the theoretically possible 138 CV syllables are actually found in the data, i.e. the language uses 99% of the possibilities offered by the system. The number of CVC syllables in the DBS is 1792, which is 56% of the theoretical maximum. But only 1% of the maximal number of possible basic CVCC syllables actually occurs in the language. There is a dramatic decrease in the possible occurences when the complexity of the syllable increases. The most simple syllable structure, CV, allows relatively few different sequences, but practically all of them are used in the language. The more complex syllable structure, CVC, allows considerably more sequences, but no more than a little over half of them are used in the language. Finally, the most complex syllable type, CVCC, allows a vast number of theoretically possible different sequences, but only one per cent of these are found in the data. One explanation for the unused possibilities (gaps) is phonotactic constraints imposed by the language, i.e. certain kinds of segment sequences are not allowed. Phonotactic constraints of Modern Persian will be discussed in detail in Chapters 5 and 6. However, phonotactic constraints do not account for all gaps of which a great number are simply accidental.

Table 3.2: *Maximal number of possible syllables for each syllable type, frequencies of each type in the DBS, and percentages of actual frequencies out of the maximal*

Syllable type	Maximum	DBS	% of maximum
CV	138	137	98.55
CVC	4416	1792	56.49
CVCC	73002	772	1.05
Total	**77556**	**2701**	**3.53**

Appendix 5 gives an etymologically coded list of basic syllables and their frequencies in the DW, in a descending order of frequency. If we take the twenty most frequent syllables from Appendix 5, and their frequencies, we get the following list:

	Syllable	Frequency		Syllable	Frequency
1.	/tä/	590	**11.**	/va/	211
2.	/mo/	584	**12.**	/re/	208
3.	/ʔa/	389	**13.**	/ʔä/	196
4.	/te/	288	**14.**	/ma/	188
5.	/de/	239	**15.**	/le/	172
6.	/mä/	235	**16.**	/sa/	158
7.	/ra/	233	**17.**	/la/	158
8.	/dän/	231	**18.**	/ba/	155
9.	/na/	229	**19.**	/sä/	147
10.	/ne/	229	**20.**	/nä/	137

As the list shows, the frequencies decrease very rapidly: no. 1 has a frequency of 590, while no 20. has a frequency of 137. According to Appendix 5, altogether 1060 basic syllables (about 40% of the DBS) have a frequency of one in the DW, and 1953 basic syllables (about 73%) have a frequency of five or less in the DW. All but one of the top twenty syllables are of the type CV. The high frequency of the exceptional syllable, /dän/, no. 8, is mainly due to the fact that, among other things, it is found in infinitives.

The distribution of the three syllable types (CV, CVC, CVCC) as a function of their ordinal number.[39]is shown in Figure 3.1. The list of basic syllables in Appendix 5 is divided into 27 groups, so that each group contains 100 syllables, except for the last group, which contains 101 syllables. The first group contains one hundred most frequent syllables, i.e. the syllables with ordinal numbers from 1 to 100, the second group contains the syllables with ordinal numbers from 101 to 200, etc. The curves in Figure 3.1 show how the one hundred syllables in each group are distributed among CV, CVC and CVCC syllables.

[39] The ordinal number is meant to reflect the frequency of the syllable: the most frequent syllable has the number 1, the syllable with the next highest frequency has the number two, etc. However, there is no one-to-one correspondence between the ordinal number of the syllable and its frequency, because several syllables can have the same frequency. In such cases the syllables are listed alphabetically, and the ordinal number is assigned according to the alphabetical order.

Figure 3.1: *Distribution of CV, CVC and CVCC syllables in the DBS*

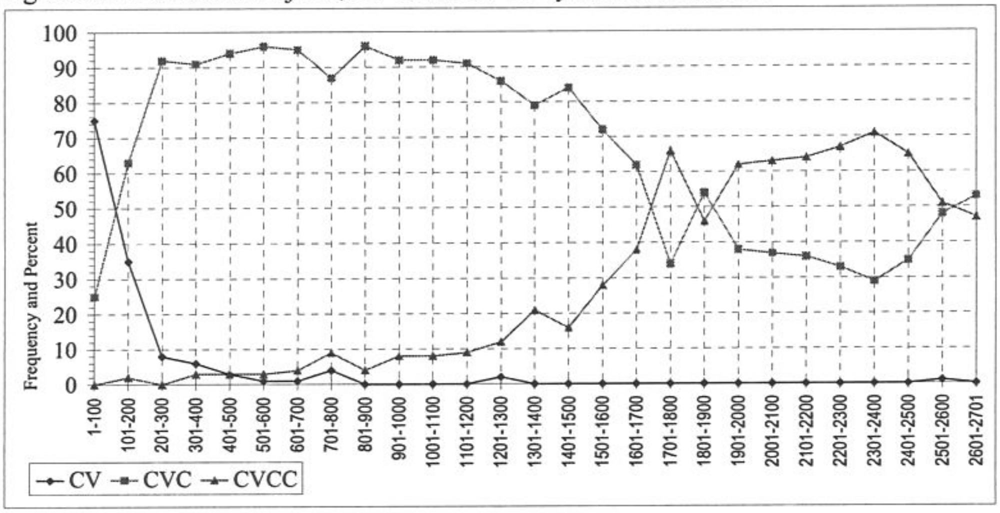


Figure 3.1 does not mention the frequencies of the 2701 syllables in the DW, but we can add some figures on the basis of Appendix 5 to show the rapid decrease of frequencies: in the first group (1-100) frequencies range from 49 to 590, in the second group (101-200) from 49 to 22, in the third group (201-300) from 22 to 14. In the middle of the eighth group (701-800), syllable frequency has gone down to five, and starting about the middle of the seventeenth group (1601-1700; starting from no. 1641), all the remaining syllables have the frequency of one in the DW. Figure 3.1 shows that of the 100 most frequent syllables, 75 are CV syllables. This is nearly half of all the 137 CV syllables in the DBS. The remaining 25 of the top one hundred are CVC syllables. In the next group, i.e. syllables with ordinal numbers 101 to 200, the number of CV syllables is reduced to 32, while the number of CVC syllables has risen to 66. Altogether 107 CV syllables, about 80% of all the basic CV syllables, are among the 200 most frequent syllables. In the interval between 300 and 1700, CVC is the dominant type, with a proportion of 90% or more in several groups, and CVCC is the second in most groups. In the interval between 1700 and 2700, where syllable frequency in the DW is one, CVCC is the dominant type in almost all groups, with a proportion of 60-70%, and CVC takes the remaining 40-30%. There are practically no CV syllables in this interval. In other words, CV syllables are

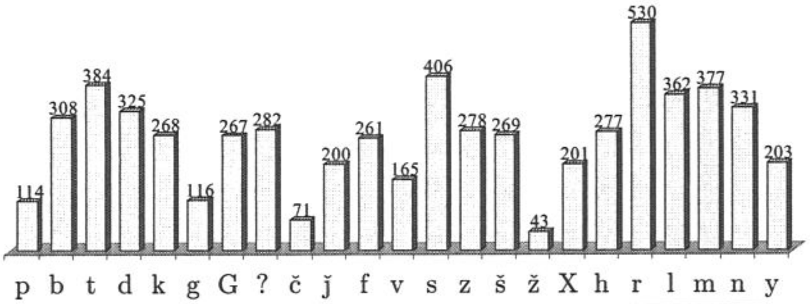
mainly high-frequency syllables, and CVCC syllables tend to have low frequencies, while CVC syllables range over the whole frequency scale.

The rest of this chapter describes the DBS on the level of segments and segment classes. Comparisons will be made between the DBS and the DW. Frequencies of segments and segment classes will also be compared between the etymologically Persian and non-Persian data. No attention is paid here to the syllable type, the analysis of which will be taken up later, in chapters 4-6.

3.2. Consonant frequencies

The syllables in the DBS contain 6038 consonant tokens. Consonant frequencies are given in Figure 3.2. The proportions of consonants range from 8.78% for /r/ to 0.71% for /ž/. Consonant frequencies are somewhat more evenly distributed in the DBS than in the DW (see Figure 2.2). Sonorants have high frequencies, and four of the five sonorants are among the top six consonants in the DBS. Of the fricatives, /s/ is the most frequent one, and dentals have the highest frequencies of stops.

Figure 3.2: *Frequencies and percentages of consonants in the DBS (DO = in Descending Order)*



p	1.89%	d	5.38%	G	4.42%	ǰ	3.31%	s	6.72%	ž	0.71%	r	8.78%	m	5.48%
b	5.10%	k	4.44%	ʔ	4.67%	f	4.32%	z	4.60%	χ	3.33%	l	6.00%	y	3.36%
t	6.36%	g	1.92%	č	1.18%	v	2.73%	š	4.46%	h	4.59%	n	6.24%		100.00%

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

Figure 3.3 compares the percentages of the consonants in the DBS and the DW. The white column means that the consonant has a bigger proportion in the

DW, and the shaded column indicates that the proportion is bigger in the DBS. The figure shows that three of the sonorants, /r, m, n/, and three stops, /t, d, ʔ/, have a higher percentage in the DW. This means that they are common (also) in higher-frequency syllables in the DW.

Figure 3.3: *Comparison of the percentages of the consonants in the DBS and DW*

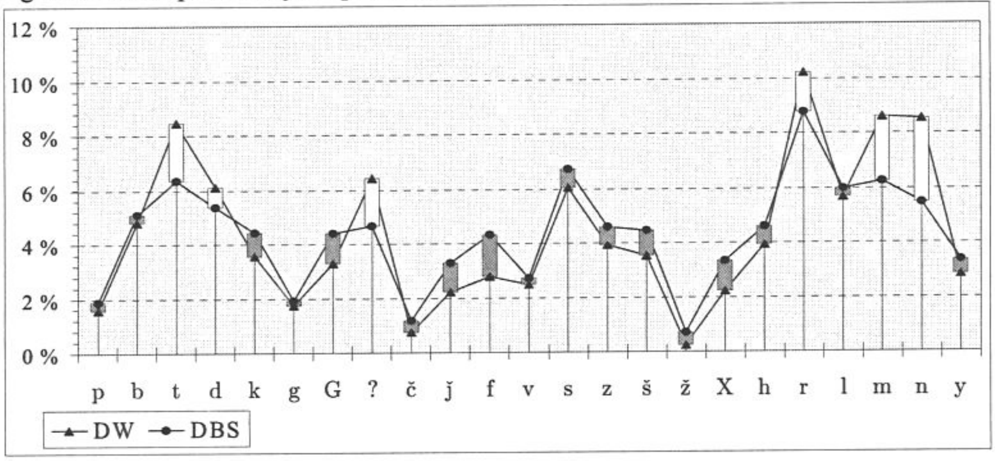
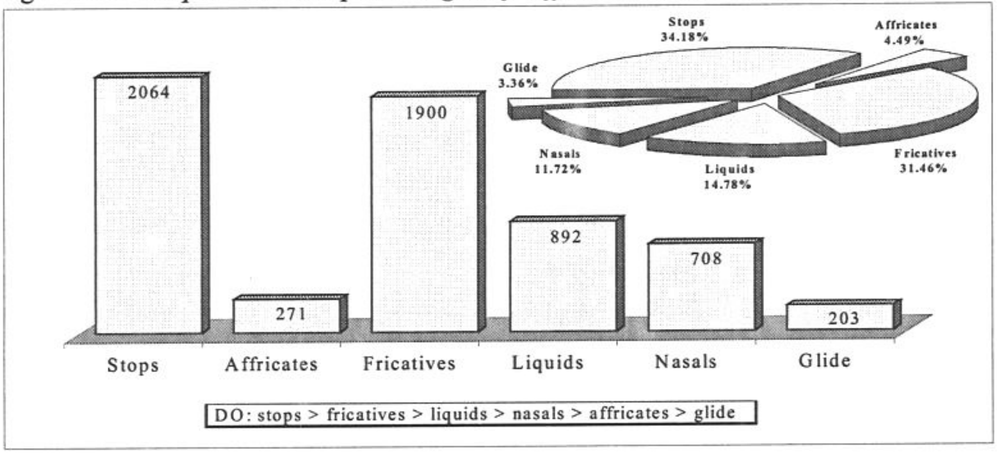


Figure 3.4 gives the frequencies and percentages of different manner subclasses in the DBS. Stops and fricatives are the most frequent subclasses, each of them taking roughly one third of the consonant tokens. Liquids cover about 15% and nasals 12%.

Figure 3.4: *Frequencies and percentages of different manner subclasses in the DBS*



The proportions of different manner subclasses in the DBS are compared to those in the DW in Figure 3.5. The white column indicates that the proportion is bigger in the DW, and the shaded column indicates that the proportion is bigger in the DBS.

Figure 3.5: *Comparison of percentages of the manner subclasses in the DBS and the DW*

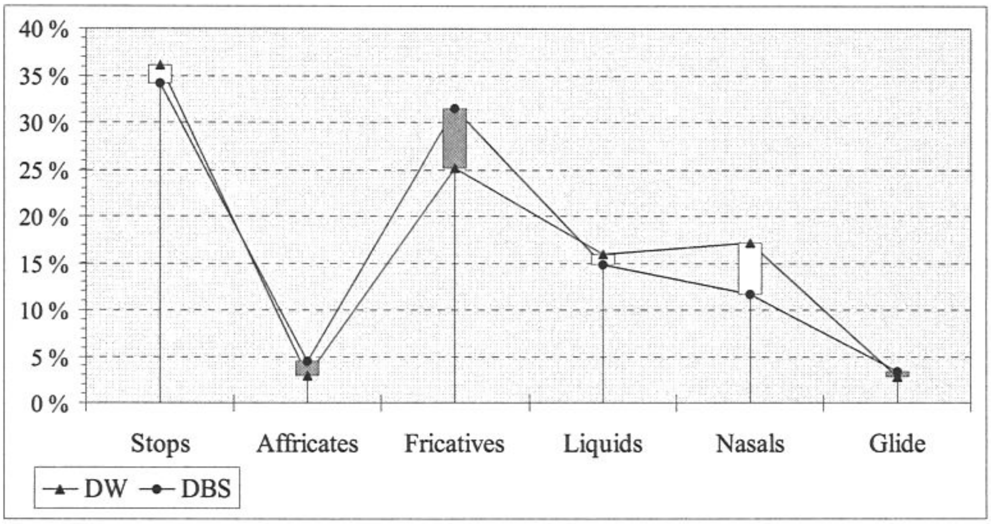


Figure 3.5 shows that nasals have a clearly higher percentage in the DW, while the proportion of fricatives is considerably bigger in the DBS. This implies that a great number of fricative tokens are found in low-frequency syllables, while higher-frequency syllables contain a considerable number of nasals.

Of the major manner classes, obstruents have a proportion of 70.13% in the DBS, and sonorants have a proportion of 29.86%. These percentages are somewhat different in the DW in which obstruents have 64.13% and sonorants 35.88% of consonant tokens. As we saw above in Figure 3.5, it is mainly fricatives and nasals that account for the fact that the major subclasses have different proportions in the two data.

Figure 3.6 presents the frequencies of place classes in the DBS. Dentals, with a proportion of 44%, are the biggest group. Labials, with a proportion of roughly 20%, come next, and velars are the smallest group.

Figure 3.6. *Frequencies and percentages of place classes in the DBS*

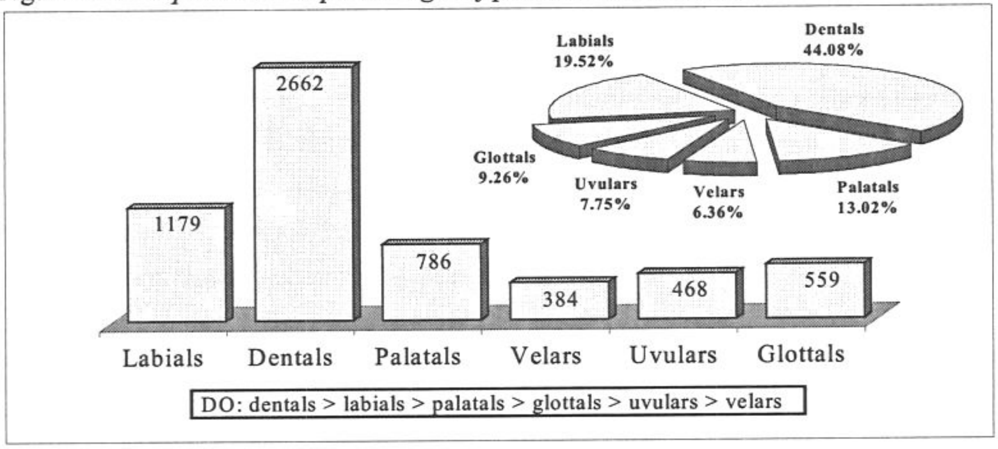
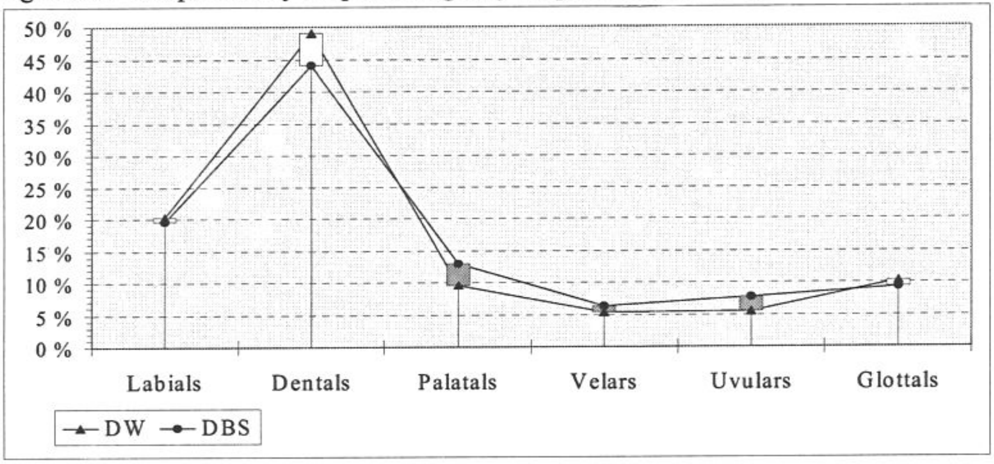


Figure 3.7 compares the percentages of place classes in the DW and DBS. A white column indicates a higher percentage in the DW, whereas a shaded column means that the percentage is higher in the DBS. The figure shows that the percentage of dentals is higher in the DW, i.e. syllables with higher frequencies favour dentals. On the other hand, palatals and also uvulars have a somewhat bigger proportion in the DBS. This means that consonants of these place classes are more common in low-frequency syllables. However, the differences are not great; the descending orders of place class frequencies in the DW and the DBS are the same with two exception (see Figures 3.4 and 2.4), namely, glottals and palatals are in opposite orders.

Figure 3.7: *Comparison of the percentages of the place classes in the DBS and the DW*



Since the frequencies of consonant classes reflect both the frequencies of their member consonants and also the number of consonants in the class, the average frequencies per consonant in each class were calculated in addition. The averages are given in Tables 3.2-3.5. We see that the average frequency of a sonorant is clearly higher than that of an obstruent, even though sonorants as a group comprise only 30% of consonant tokens. Table 3.4 shows that two sonorant subgroups come first: liquids have the highest average frequency and nasals the next highest. After them come stops and fricatives. Finally, Table 3.5 indicates that among the place classes, dentals have the highest average frequency. In other words, dentals are the most common place class, no matter whether it is measured by average frequencies per consonant or by frequencies of consonant tokens in the place class. The second group in average frequencies is glottals – they are fourth when measured by the frequency of all tokens in the class. The third highest average belongs to labials.

Table 3.2. *Average frequencies of manner classes in the DBS*

Obstruents	Sonorants
235.2	360.6

Table 3.3. *Average frequenies of voiced vs. voiceless obstruents in the DBS*

Voiceless obstruents	Voiced obstruents
253.3	212.75

Table 3.4. *Average frequencies of manner subclasses in the DBS*

Stops	Affricates	Fricatives	Liquids	Nasals	Glide
258	135.5	237.5	446	354	203

Table 3.5. *Average frequencies of place classes in the DBS*

Labials	Dentals	Palatals	Velars	Uvulars	Glottals
235.8	380.2	157.2	192	234	279.5

3.3. Vowel frequencies

Figure 3.8 presents vowel frequencies and percentages in the DBS. We see a similar phenomenon in vowels as we saw in consonants in the DBS, frequencies are more evenly distributed in the DBS than in the DW. The percentages of vowels range from 28% for /ä/ to 10% for /u/. (The corresponding figures in the DW are 31% for /ä/ and 5% for /u/, see Figure 2.5.) Close vowels have the lowest frequencies, the open vowel /ä/ has the highest frequency, and the rest – mid vowels and the other open vowel /a/ – have frequencies in between. Thus, there is a correlation between vowel frequency and sonority, but it is weakened by the fact that /a/ is an exception, resembling a mid vowel in terms of frequencies. However, instead of treating the vowels as one group, we can also treat them as two separate groups consisting of short (/ä, e, o/) and long (/a, i, u/) vowels. There are no exceptions now: in each group, the frequency decreases as the sonority decreases.

Figure 3.8: *Frequencies and percentages of vowels in the DBS*

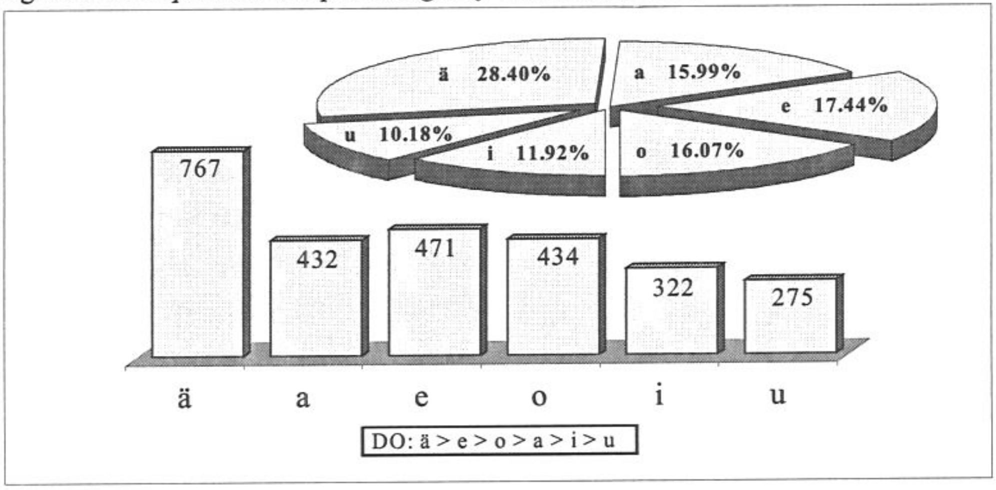


Figure 3.9 compares the percentages of the vowels in the DBS and the DW. A white column means that the vowel has a bigger proportion in the DW, while a shaded column indicates that the proportion of the vowel is bigger in the DBS. The figure shows that the top three vowels, /a, ä, e/ and especially /a/, have a bigger proportion in the DW. In other words, they are more common in higher-frequency syllables. The rest, /o, i, u/, have bigger proportions in the DBS. This means that they occur more in lower-frequency syllables.

Figure 3.9: *Comparison of percentages of the vowels in the DW and the DBS*

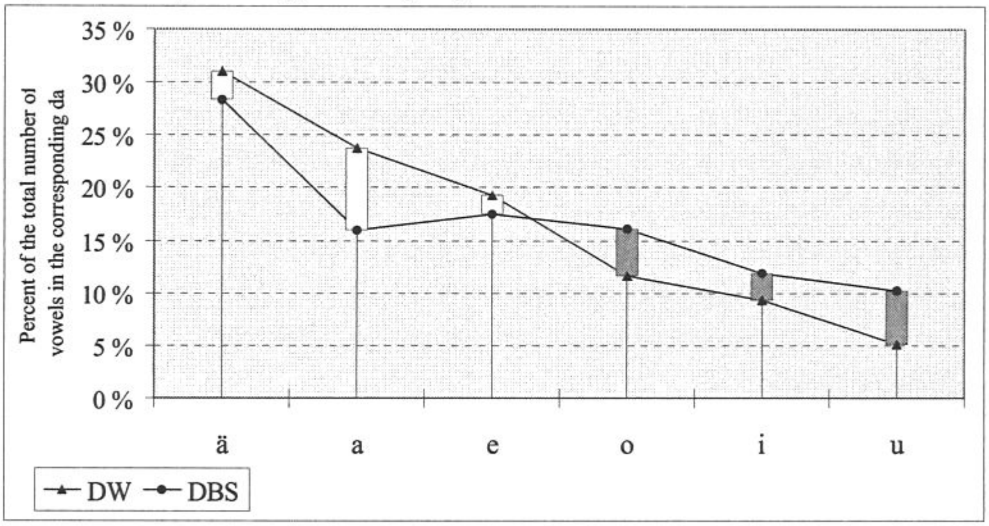
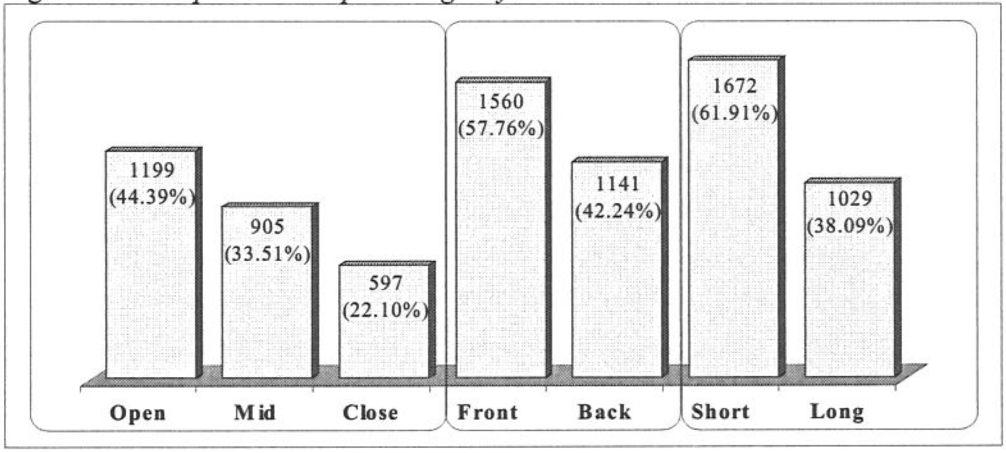


Figure 3.10 gives frequencies of vowel classes and their percentages of vowel tokens in the DBS. The figure shows that open vowels have the biggest percentage and close vowels have the smallest percentage, that front vowels are more frequent than back vowels. and that short vowels are more frequent than long vowels.

Figure 3.10. *Frequencies and percentages of vowel classes in the DBS*



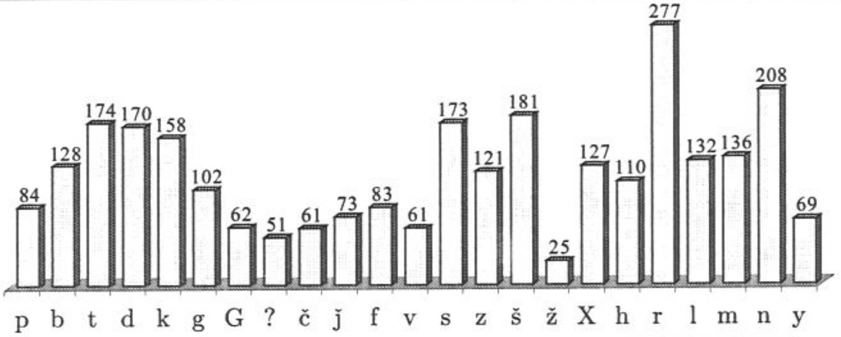
3.4. Phoneme frequencies and etymology

3.4.1. Consonants

DPS

It was mentioned before that the data of etymologically Persian syllables (DPS) consists of 1315 syllables. They contain 2766 consonant tokens, which is 45.81% of the total number of consonants in the DBS. Figure 3.11 gives consonant frequencies in the DPS, together with their percentages. The figure shows that the top two are sonorants, namely, /r/ and /n/. The highest frequencies in the fricative group belong to /š/ and /s/, and the most frequent stops in the DPS are the two dentals and the voiceless velar /k/.

Figure 3.11: *Frequencies and percentages of consonants in the DPS*

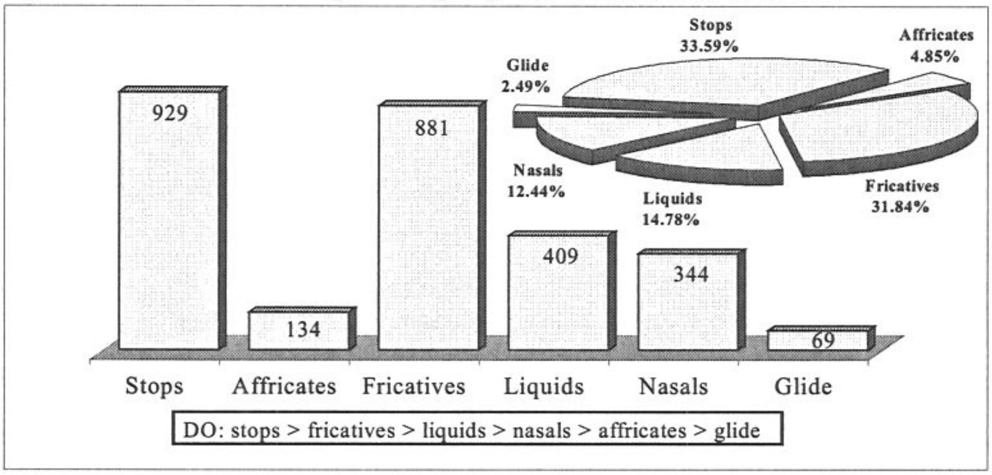


p	3.04%	d	6.15%	G	2.24%	ǰ	2.64%	s	6.25%	ǰ	0.90%	r	10.01%	n	7.52%
b	4.63%	k	5.71%	ʔ	1.84%	f	3.00%	z	4.37%	χ	4.59%	l	4.77%	y	2.49%
t	6.29%	g	3.69%	č	2.21%	v	2.21%	š	6.54%	h	3.98%	m	4.92%		100.00%

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

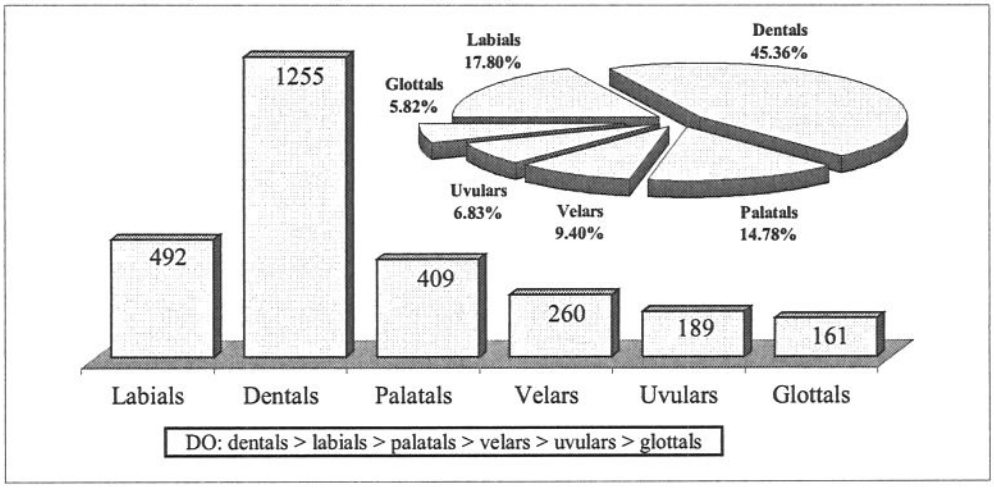
Figure 3.12 shows frequencies of manner subclasses and their percentages in the DPS. The biggest groups are stops (34%) and fricatives (32%). The proportions of all manner subclasses are practically the same as in the DBS (see Figure 3.4).

Figure 3.12: *Frequencies and percentages of manner subclasses in the DPS*



The frequencies and percentages of different place classes are given in Figure 3.13. The figure shows that dentals, with a proportion of 45%, are by far the biggest group, while the percentages of the rest range from 18% (labials) to 6% (glottals). There are some (but not great) differences between the DPS and the DBS in the percentages of the lower-frequency place classes, e.g. velars have a bigger percentage in the DPS (9%) than in the DBS (6%), while the case of the glottals is the opposite (6% in the DPS; 9% in the DBS).

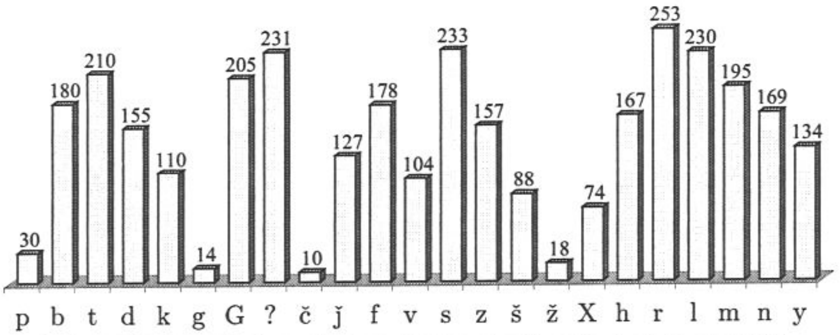
Figure 3.13: *Frequencies and percentages of place classes in the DPS*



DNPS

It was mentioned before that the data of non-Persian origin (DNPS) consists of 1386 syllables. They contain 3272 consonant tokens, which is 54.19% of all consonant tokens in the DBS. Figure 3.14 presents the frequencies and percentages of the consonants in the DNPS. The figure shows that the most frequent consonant is /r/, as it is in the DPS and DBS, but there are some others that are nearly as frequent, i.e. /1/, /s/ and /ʔ/. On the other hand, four consonants are quite uncommon in this data, namely /p/, /g/, /č/, and /ž/, each having a proportion of less than 1%.

Figure 3.14: *Frequencies and percentages of consonants in the DNPS*

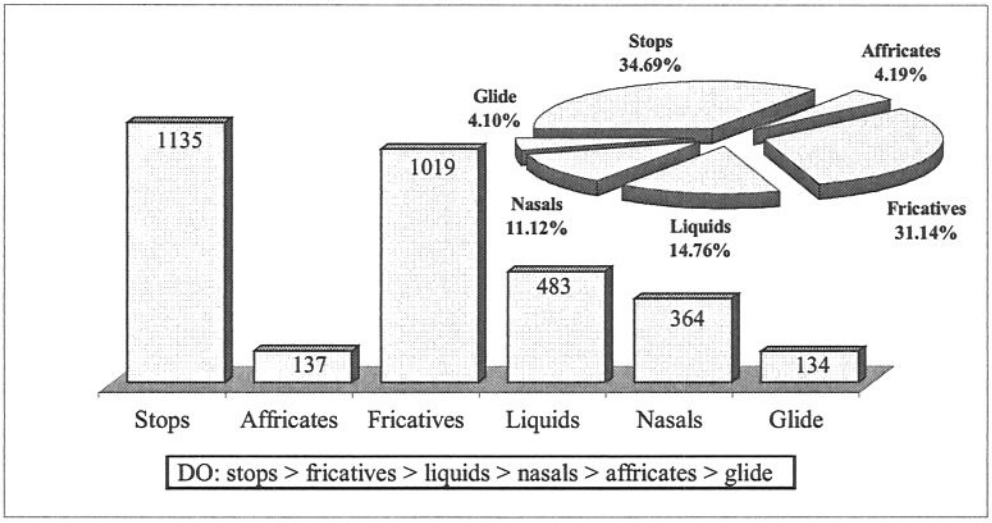


p	0.92%	d	4.74%	G	6.27%	ǰ	3.88%	s	7.12%	ž	0.55%	r	7.73%	n	5.17%
b	5.50%	k	3.36%	ʔ	7.06%	f	5.44%	z	4.80%	χ	2.26%	l	7.03%	y	4.10%
t	6.42%	g	0.43%	č	0.31%	v	3.18%	š	2.69%	h	5.10%	m	5.96%		100.00%

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

Figure 3.15 displays frequencies of manner subclasses and their percentages in the DNPS. We see that stops (35%) and fricatives (31%) are the biggest groups. The percentages of the rest range from 15% (liquids) to 4% (affricates and the glide each).

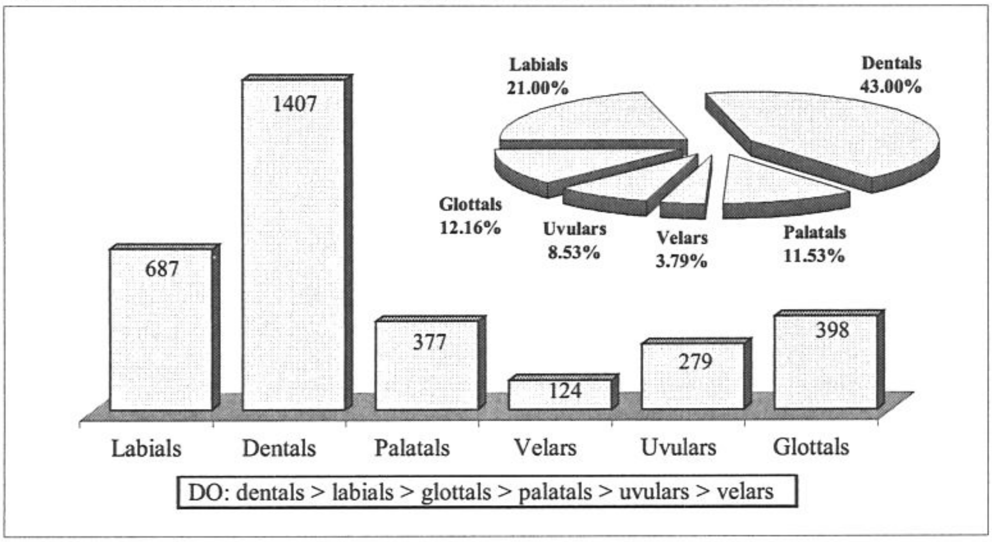
Figure 3.15: *Frequencies and percentages of manner subclasses in the DNPS*



DO: stops > fricatives > liquids > nasals > affricates > glide

Figure 3.16 presents frequencies and percentages of place classes in the DNPS. Dentals are the biggest group, with a proportion of 43%, and labials are the next biggest group (21%). The percentages of the rest range from 12% (glottals and palatals each) to 4% (velars).

Figure 3.16: *Frequencies and percentages of place classes in the DNPS*

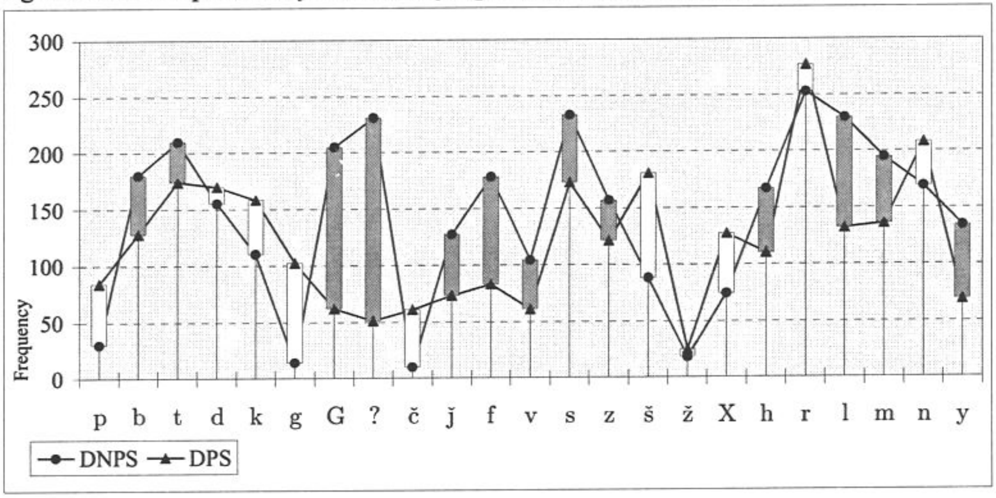


DO: dentals > labials > glottals > palatals > uvulars > velars

Comparison

Figure 3.17 compares consonant frequencies in the DBS and DNPS. A shaded column means that the consonant frequency is higher in the DNPS and the white column indicates that the frequency is higher in the DPS. We can recall here that 54% of consonant tokens in the DBS come from the DNPS and 46% from the DPS. If frequencies were distributed evenly we would expect these proportions to hold for each consonant, i.e. we would expect only shaded columns (and rather small ones) in the picture. However, this is not the case, as the figure shows. Ten consonants have a white column, i.e. their frequencies are higher in the DPS. In some of them, e.g. /g/ and /č/, almost all tokens come from the DPS. In the rest of the consonants, tokens of non-Persian origin are in the majority (shaded columns), but the percentages of tokens can be considerably more than 54%, as e.g. in /G/ and /ʔ/.

Figure 3.17: *Comparison of consonant frequencies in the DPS and DNPS*



We can see the percentages of DPS vs. DNPS in Figure 3.18. It shows that only /t/ fits the ratio 54%-46%, and /b/ and /m/ come close to it. At one extreme, over 85% of the tokens of /g/ and /č/ and 74% of the tokens of /p/ come from etymologically Persian data. At the other end, more than 76% of the tokens of /G/ and /ʔ/ come from loanwords.

Figure 3.18: *Comparison of consonant tokens in the DPS and DNPS*

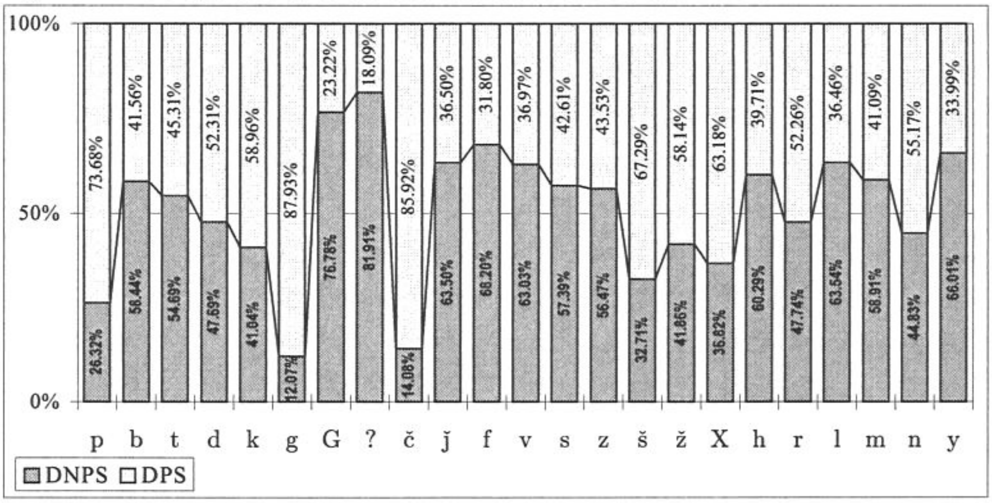


Figure 3.19 compares frequencies of manner subclasses in the DPS and the DNPS. It shows that, except for affricates, the frequencies of all other manner subgroups are higher in the DNPS.

Figure 3.19: *Comparison of frequencies of the manner subclasses in the DPS and the DNPS*

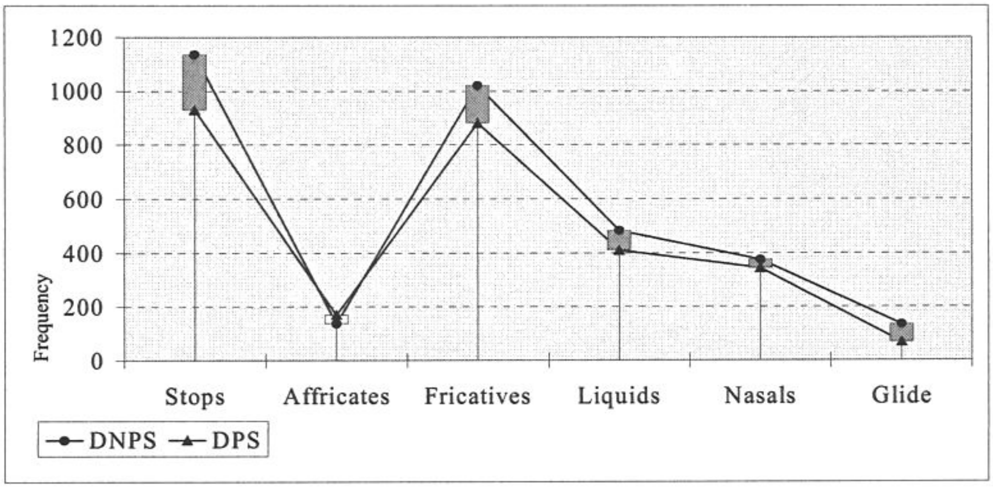


Figure 3.20 gives the percentages of consonant tokens in different manner subclasses coming from the DPS vs. the DNPS. The figure shows that the proportions correspond very well to the ratio 54%-46% in the three biggest subclasses (stops, fricatives, liquids). Moreover, the figure suggests that the lower the frequency of the class, the more the proportions deviate from the ratio. Thus, in the least frequent group, the glide, the ratio is 66% (DNPS) - 34% (DPS).

Figure 3.21 shows the percentages of DPS vs. DNPS tokens in major manner classes. We see that the ratio 54%-46% holds perfectly for both classes. The figure also shows that voiced and voiceless obstruents come close to the ratio.

Figure 3.20: *Percentages of manner subclasses in DPS and DNPS*

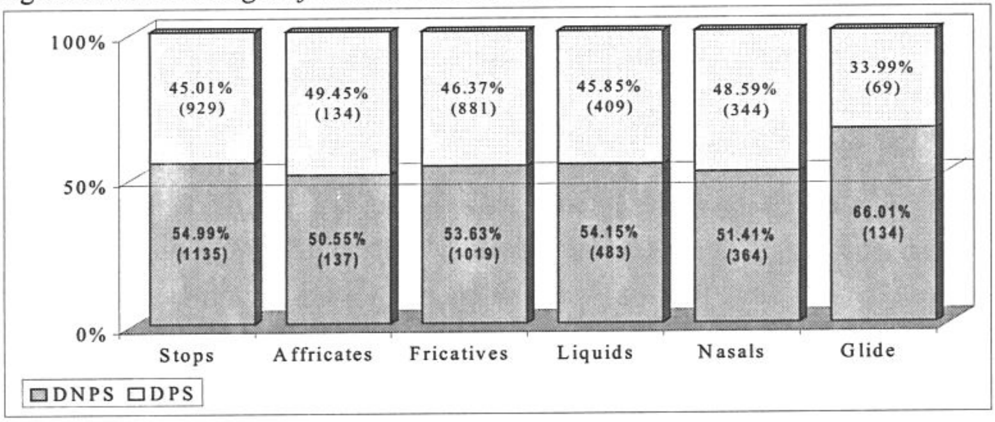


Figure 3.21: *Percentages of consonant tokens of Persian vs. non-Persian origin of manner classes and voicless/voiced obstruents in the DBS*

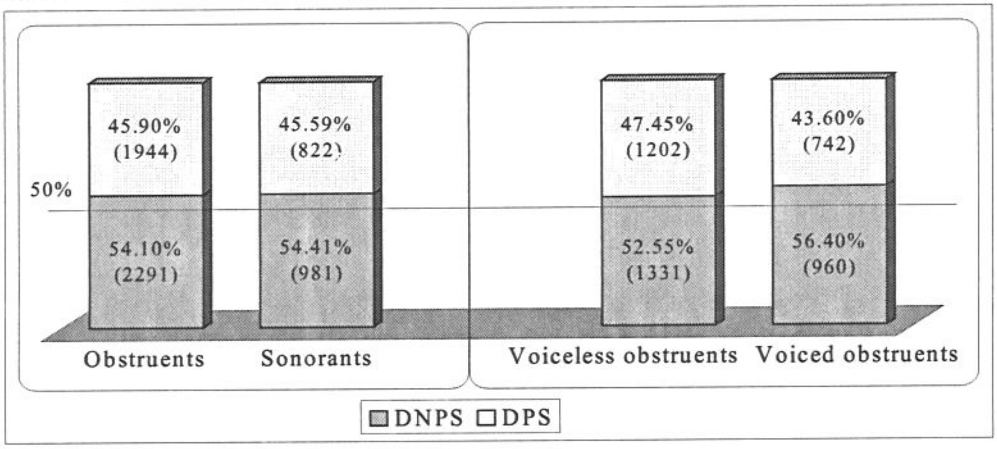


Figure 3.22 compares the DPS and the DNPS with respect to the frequencies of different place classes. The figure shows that the frequencies in the DNPS are higher in all but two classes, velars and palatals.

Figure 3.22: *Comparison of frequencies of the place classes in the DPS and the DNPS*

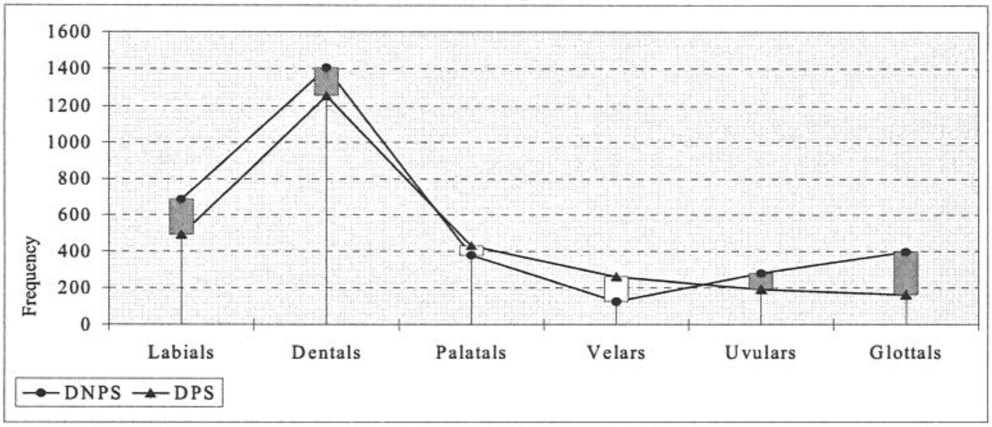
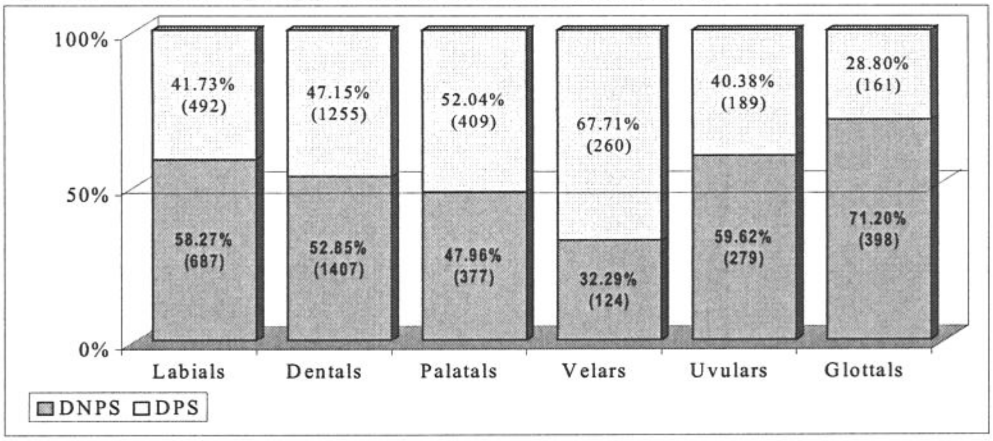


Figure 3.23 shows the percentages of place classes in the DPS and DNPS. The biggest group, dentals, comes close to the ratio 54%-46%. All the other groups deviate from it more or less. The biggest deviations are found in glottals, where the majority of tokens come from loanwords (71% DNPS - 29% DPS), and velars, where the majority of tokens comes from etymologically Persian words (32% DNPS - 68% DPS).

Figure 3.23: *Percentages of place classes in the DPS and DNPS*



It is interesting to see that the big consonant classes both in the DPS and the DNPS conform to the 46%-54% ratio, even though there are deviations in the smaller classes. A great amount of deviation from the average 54%-46% is seen on the level of individual consonants. Manner subclasses show only small deviations, and major manner classes fit the ratio perfectly. Only some place classes, such as velars and glottals, show clear deviations from the ratio, but these are small (with two members only in each).

3.4.2. Vowels

We have seen that the DBS consists of 2701 syllables, of which 1315 (48.69%) syllables come from the DPS and 1386 syllables (51.31%) come from the DNPS. Since each syllable has one vowel, the figures also represent the numbers of vowel tokens in each type of data. Figure 3.24 gives the frequencies and percentages of vowels in the DPS and the DNPS. The figure shows that frequencies are more evenly distributed in the DPS, with the percentage of vowels ranging from 27% to 12%; the corresponding figures in the DNPS are 30% and 9%. In the DPS, open vowels have the highest frequencies, and close vowels have the lowest frequencies. Thus, there is a high positive correlation between vowel frequency and sonority. The DNPS shows a similar tendency, but /a/ is an exception having the second lowest frequency, i.e. it is like a close vowel in terms of frequencies. We now see that the exceptionally low frequency of /a/ in the Data of Basic Syllables is due to the small proportion of this vowel in loanwords.

Figure 3.24: *Frequencies and percentages of the vowels in the DPS and the DNPS*

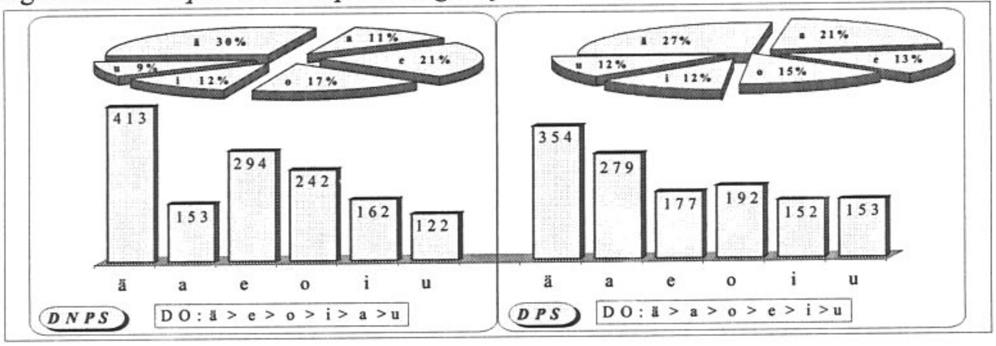


Figure 3.25 compares frequencies of vowel classes in the DPS and the DNPS. A white column indicates that the vowel is more frequent in the DPS, and a shaded column shows that the vowel is more frequent in the DNPS. Since 51% of vowel tokens come from the DNPS, we could expect short shaded columns for all vowels, provided that the ratio 51%-49% shows up for each vowel. However, the figure shows that two vowels in particular, /a/ and /e/, deviate clearly from the average: /a/ has a lower frequency in the DNPS and /e/ has a higher frequency. The figure also shows that /u/ is more frequent in the DPS than the average percentage (49%) presupposes.

Figure 3.25: *Comparison of vowel frequencies in the DPS and the DNPS*

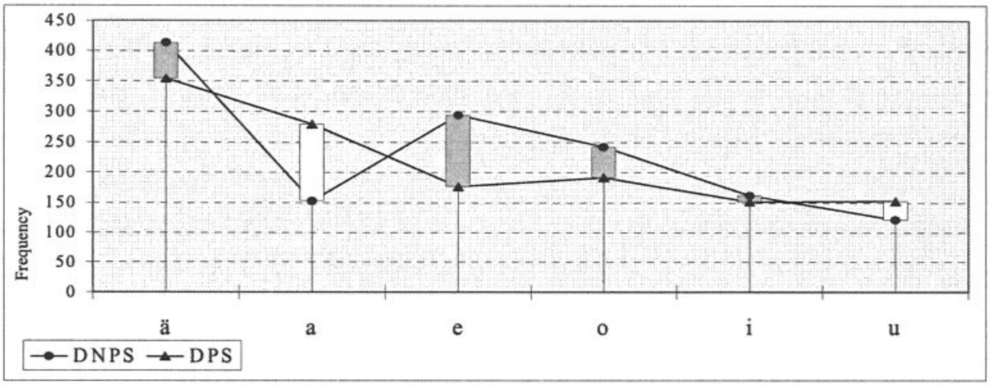
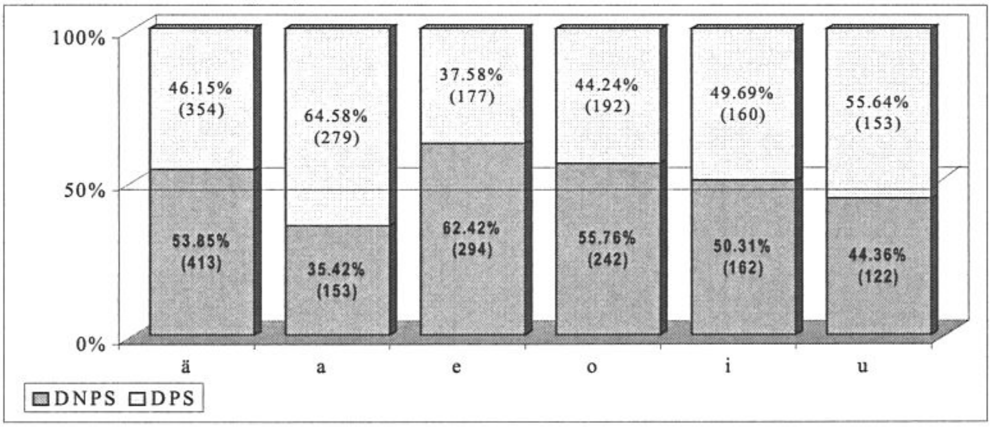


Figure 3.26 shows what percentages of the tokens of each vowel come from the DPS vs. the DNPS.

Figure 3.26: *Percentages of vowel tokens in the DPS and DNPS*



According to the Figure 3.26, only /i/ comes close to the ratio 51% of DNPS tokens - 49% of DPS tokens. The greatest deviations from the ratio are seen in the proportions of /a/ (DNPS 35% - DPS 65%) and /e/ (DNPS 62% - DPS 38%).

Frequencies of vowel classes in the DPS vs. the DNPS are compared in Figure 3.27. A shaded column indicates that the frequency is higher in the DNPS. The exceptionally low frequency of /a/ and the rather high frequencies of /e/ and /ä/ in the DNPS, as well as the relatively high frequency of /u/ in the DPS, are clearly reflected in the vowel groups; only mid vowels have a higher frequency in the DNPS, while both open and close vowels are more frequent in the DPS. Front vowels are more frequent in the DNPS, back vowels in the DPS. Finally, short vowels (/ä, e, i/) are more frequent in the DNPS, and long vowels in the DNPS. The difference between the frequencies of short vs. long vowels is particularly big in the DNPS, where the number of short vowel tokens is more than twice the number of long vowel tokens.

Figure 3.27: *Comparison of the frequencies of vowel classes in the DPS and the DNPS*

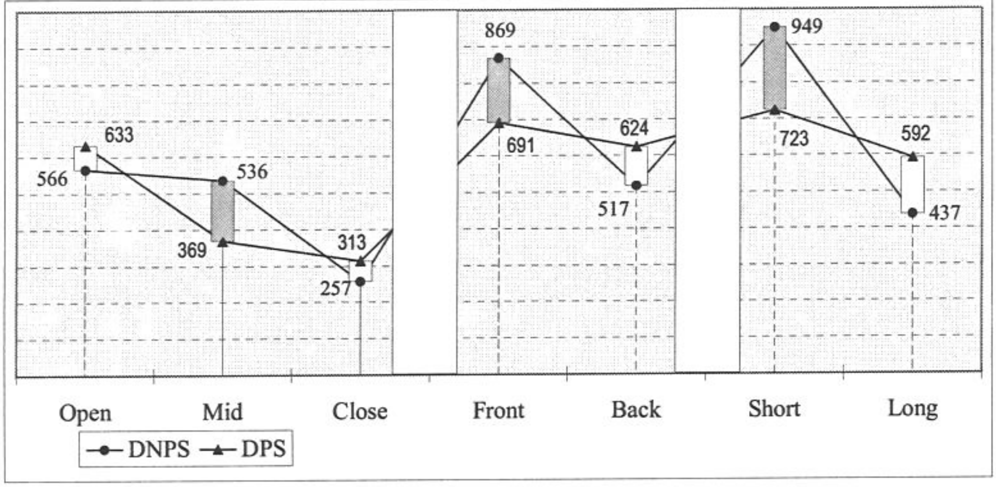
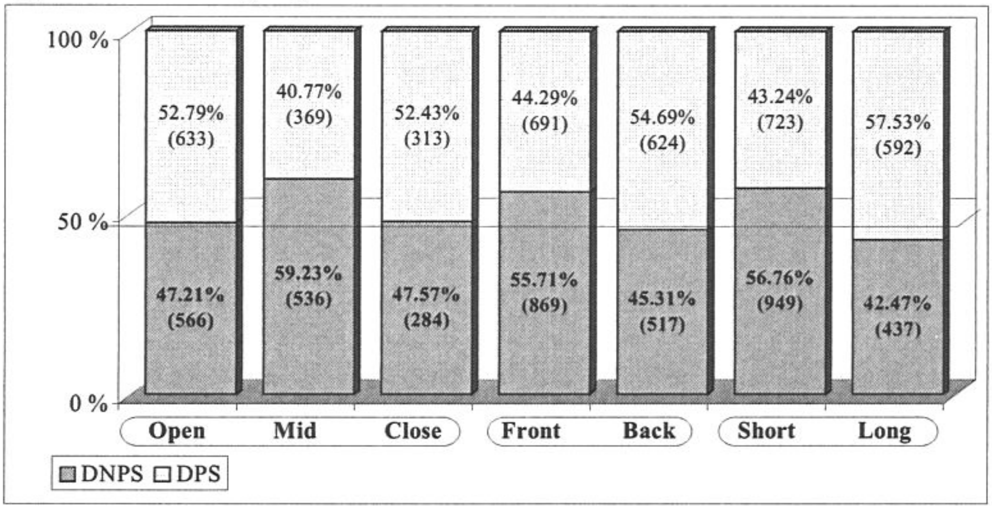


Figure 3.28 shows the percentage of tokens in various vowel classes coming from the DPS vs. the DNPS. The figure shows that all groups deviate more or less from the average 51% of DNPS tokens and 49% of DPS tokens. The greatest deviations are found in the class of mid vowels (59% of DNPS tokens - 41% of DPS tokens) and in long vowels (42% of DNPS tokens - 58% of DPS tokens).

Figure 3.28: *Percentages of vowel classes in the DPS and DNPS*


52.79% (633)
47.21% (566)
40.77% (369)
59.23% (536)
52.43% (313)
47.57% (284)
44.29% (691)
55.71% (869)
54.69% (624)
45.31% (517)
43.24% (723)
56.76% (949)
57.53% (592)
42.47% (437)
Open
Mid
Close
Front
Back
Short
Long
▨ DNPS □ DPS