5. CVCC syllables in the DBS

It was mentioned before that the Data of Basic Syllables contains 772 syllables of the type CVCC, which is only 1% of the theoretically possible CVCC syllables that occur in Persian.

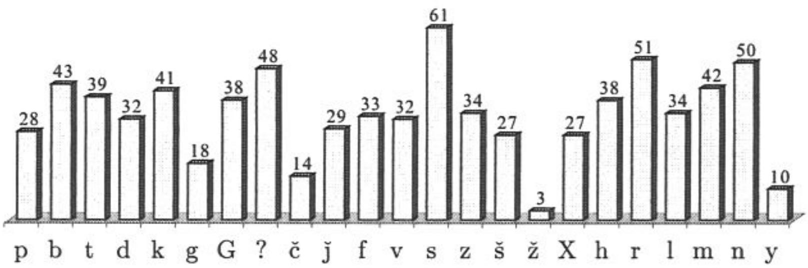
I shall begin this chapter by describing occurrences of segments in CVCC syllables in the DBS. This is done by first dealing with each position separately. Then I shall compare segments and segment classes in different syllabic positions. I shall also see what restrictions if any there are on segment combinations. CC sequences in the coda are especially on the focus.

5.1. Onset

Figure 5.1 (next page) presents consonant frequencies in the onset of CVCC syllables in the DBS. The figure shows that all consonants can occur here, but only in a few _VCC combinations.

The theoretically maximal frequency for each onset consonant is 3174 (= 6x23x23). Thus, the most frequent onset consonant, /s/, with a frequency of 61, is found only in about 2% of the theoretically possible combinations. The overall profile of onset consonants in CVCC resembles that of CVC syllables. One clear difference is that the glide, /y/, has a considerably low frequency in CVCC syllables.

Figure 5.1: *Frequencies and percentages of the consonants in the onset consonants of the CVCC syllables of the DBS*

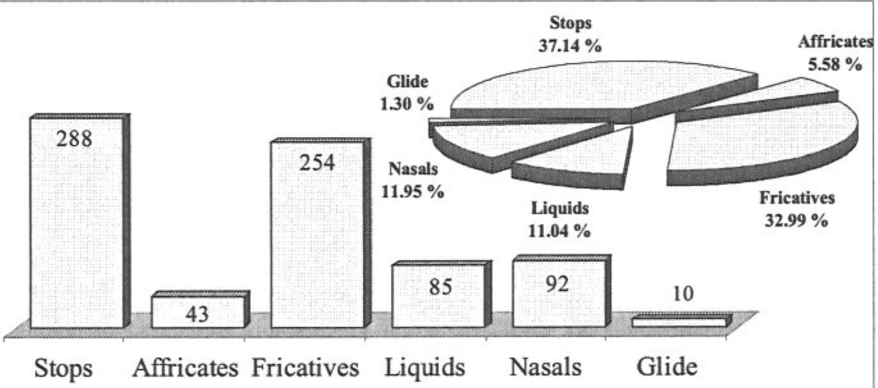


p	3.63%	d	4.15%	G	4.92%	ǰ	3.76%	s	7.90%	ž	0.39%	r	6.61%	n	6.48%
b	5.57%	k	5.31%	ʔ	6.22%	f	4.27%	z	4.40%	χ	3.50%	l	4.40%	y	1.30%
t	5.05%	g	2.33%	č	1.81%	v	4.15%	š	3.50%	h	4.92%	m	5.44%	100.00%	

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

Figure 5.2 shows frequencies and percentages of different manner subclasses. Stops (37%) and fricatives (33%) are the biggest groups. Obstruents cover about 75% of onset occurrences, sonorants about 25%. The proportions of manner classes and subclasses are similar to those of CVC syllables, except for the one member subclass of glide, which is even smaller in CVCC (1.3% vs. 4% in CVC).

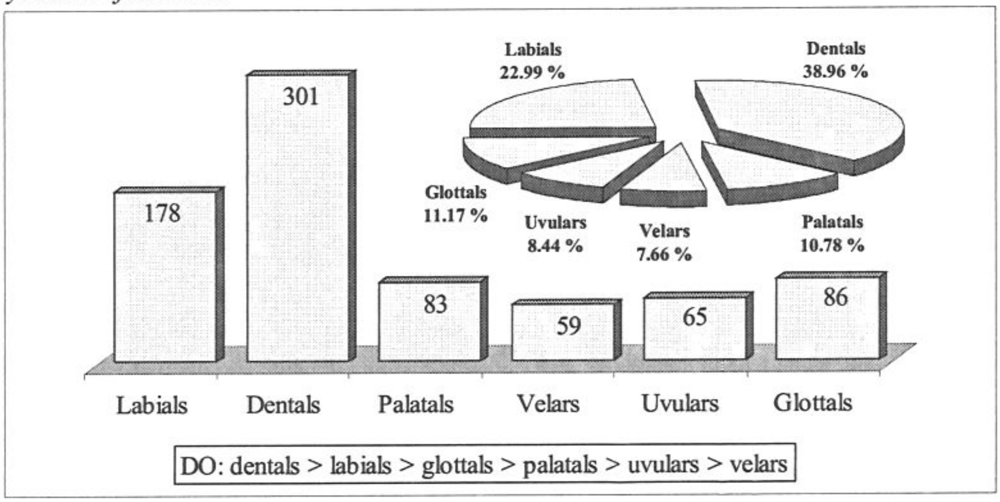
Figure 5.2: *Frequencies and percentages of manner subclasses in the onset of the CVCC syllables of the DBS*



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

Figure 5.3 gives frequencies and percentages of different place classes. Dentals (39%) is the biggest group, labials (23%) the second. The proportions of place classes in CVCC syllables are similar to those in CVC syllables, except that palatals have a lower percentage in CVCC. This is partly due to the lower frequency of the glide, but palatal affricates and fricatives are also less frequent in the onset of CVCC syllables.

Figure 5.3: *Frequencies and percentages of place classes in the onset of the CVCC syllables of the DBS*



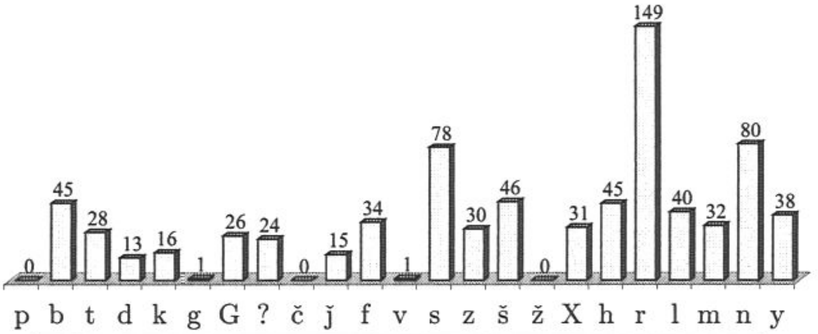
5.2. Coda

The coda contains two consonants. I shall call the first consonant in the sequence C_1, and the second consonant C_2, and I shall describe these positions separately.

5.2.1. Position C_1

Figure 5.4 shows consonant frequencies and percentages of the first coda consonant, C_1, in CVCC syllables of the DBS. We see that three of the consonants, /p/, /č/ and /ž/ do not occur at all in this position, and that two consonants, /g/ and /v/, occur only once. Three out of these five consonants are non-continuants, i.e. stops and/or affricates, and two are fricatives. In general, non-continuants tend to have low frequencies here, while some continuants, such as the sonorants /r/ and also /n/, and the fricative /s/, are very frequent. The consonant /r/ alone covers almost 20% of all C_1 occurrences .

Figure 5.4: *Frequencies and percentages of consonants in the C_1 position of the CVCC syllables of the DBS*

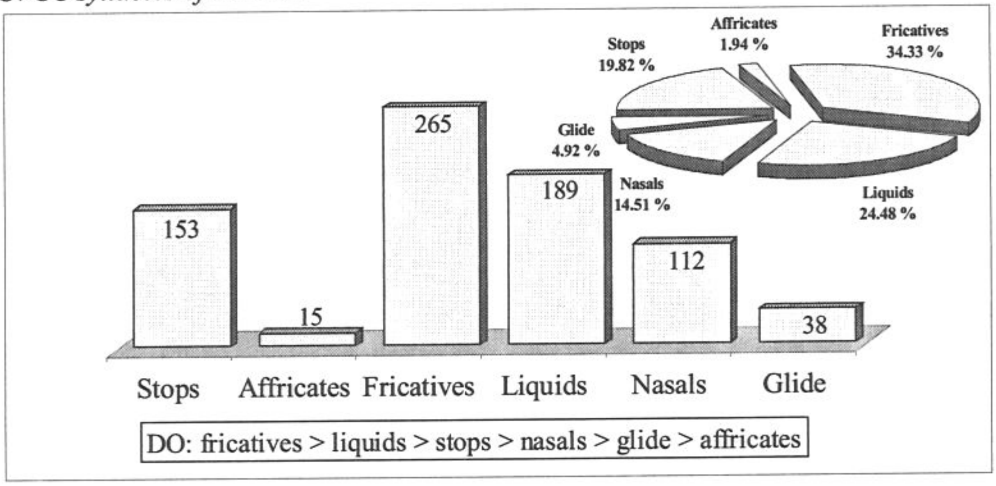


p	0.00%	d	1.68%	G	3.37%	ǰ	1.94%	s	10.10%	ž	0.00%	r	19.30%	n	10.36%
b	5.83%	k	2.07%	ʔ	3.11%	f	4.40%	z	3.89%	χ	4.02%	l	5.18%	y	4.92%
t	3.63%	g	0.13%	č	0.00%	v	0.13%	š	5.96%	h	5.83%	m	4.15%	100.00%	

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

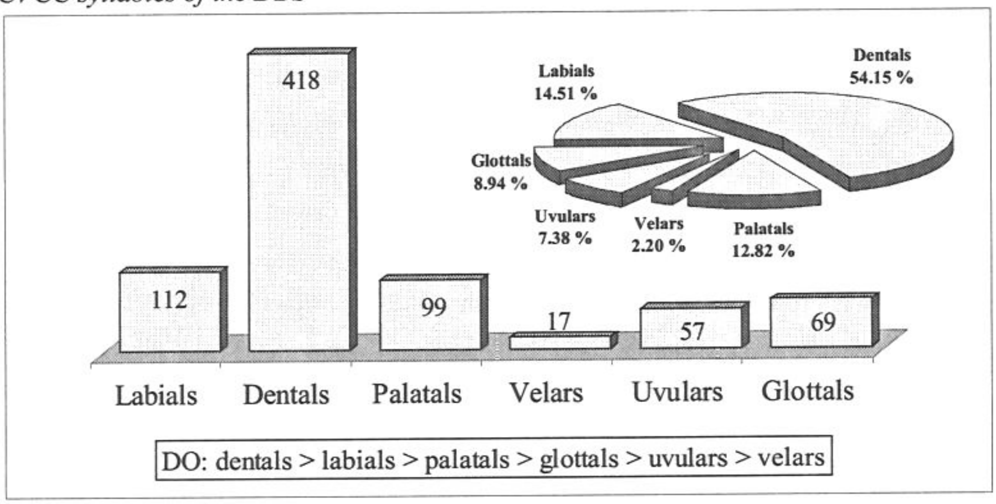
The frequencies and percentages could expect on the basis of segmental frequencies: continuants dominate, with fricatives as the biggest group (about one-third of all occurrences) and liquids the second biggest (about one-fourth). Altogether, continuants cover nearly 80% of all C_1 occurrences. These results are in harmony with the sonority hierarchy; since C_1 is not on the edge of the syllable, it is expected that this position favours more sonorous consonants, such as sonorants and fricatives, and disfavours less sonorant consonants, such as stops and affricates. Moreover, a majority (60%) of the non-continuants in this position are voiced, i.e. more sonorous than their voiceless counterparts.

Figure 5.5: *Frequencies and percentages of manner subclasses in the C_1 position of the CVCC syllables of the DBS*



The proportions of different place classes are given in Figure 5.6. As can be expected on the basis of consonant frequencies in C_1, dentals predominate with 54% of all occurrences. Since velars consist of stops only, it is to be expected that their proportion is small in C_1.

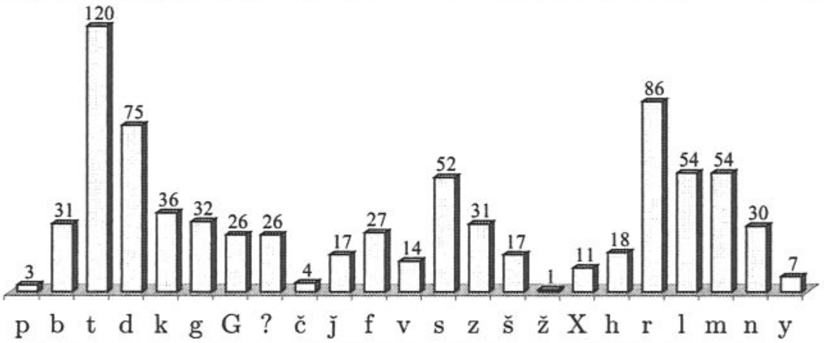
Figure 5.6: *Frequencies and percentages of place classes in the C_1 position of the CVCC syllables of the DBS*



5.2.2. Position C_2

Figure 5.7 shows consonant frequencies and percentages in C_2. All consonants can occur in C_2, but some are quite rare, e.g. /ž/, /p/, /č/ and /y/.

Figure 5.7: *Frequencies and percentages of consonants in the C_2 position of the CVCC syllables of the DBS*



p	0.39%	d	9.72%	G	3.37%	ǰ	2.20%	s	6.74%	ž	0.13%	r	11.14%	n	3.89%
b	4.02%	k	4.66%	?	3.37%	f	3.50%	z	4.02%	χ	1.42%	l	6.99%	y	0.91%
t	15.54%	g	4.15%	č	0.52%	v	1.81%	š	2.20%	h	2.33%	m	6.99%		100.00%

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

We see in Figure 5.7 three peaks; stops, especially dental stops (/t/ and /d/), have high frequencies, dental fricatives have higher frequencies than other fricatives, and the liquids – which are also dental – and /m/ are the most frequent sonorants.

Figure 5.8 presents frequencies and percentages of different manner subclasses in C_2. It is no surprise, after viewing Figure 5.7, that stops are the biggest class, with a share of 45%. Fricatives are the second biggest, with a share of 22%. The smallest group is the glide, with a share of less than 1%.

The results seen in Figure 5.8 are in harmony with the sonority hierarchy, in that sonority decreases towards the edges of the syllable. We see a clear increase in the proportions of less sonorous consonants, i.e. stops and affricates, in the C_2 position compared to the C_1 position: in C_2, stops and affricates cover about 48% of consonant occurrences, while the corresponding figure in C_1 is about 22%.

Figure 5.8: *Frequencies and percentages of manner subclasses in the* C_2 *position of the CVCC syllables of the DBS*

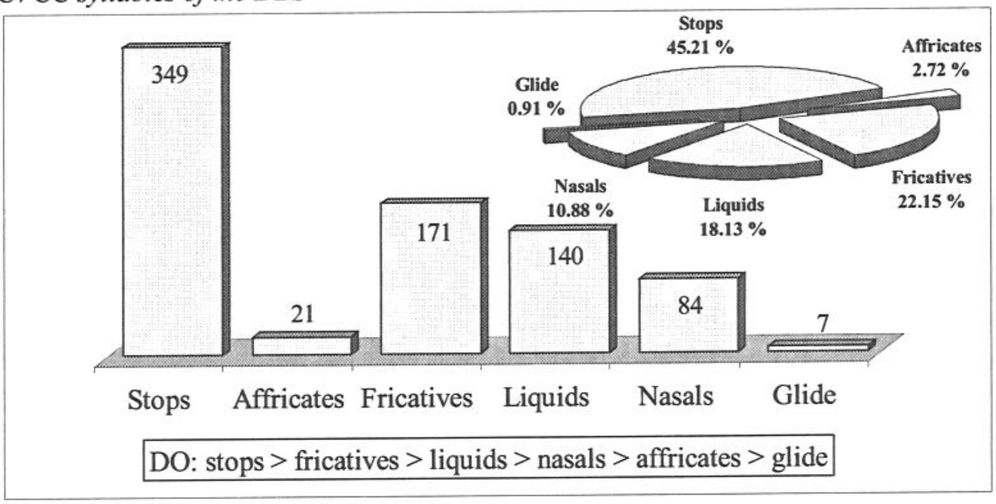
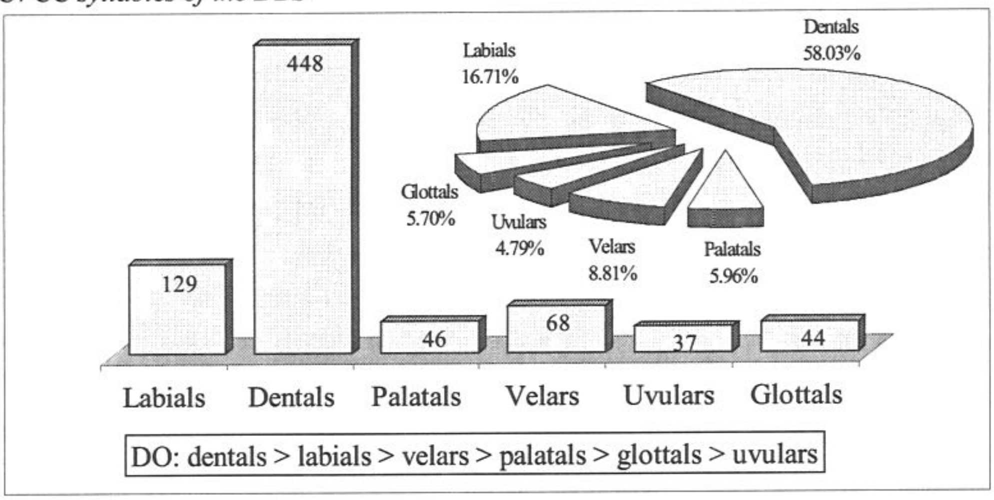


Figure 5.9 gives frequencies and percentages of different place classes in C_2. Dentals are by far the most frequent group, with a proportion of 58%. All the other groups have a proportion less than 10%, except labials, which cover about 17%.

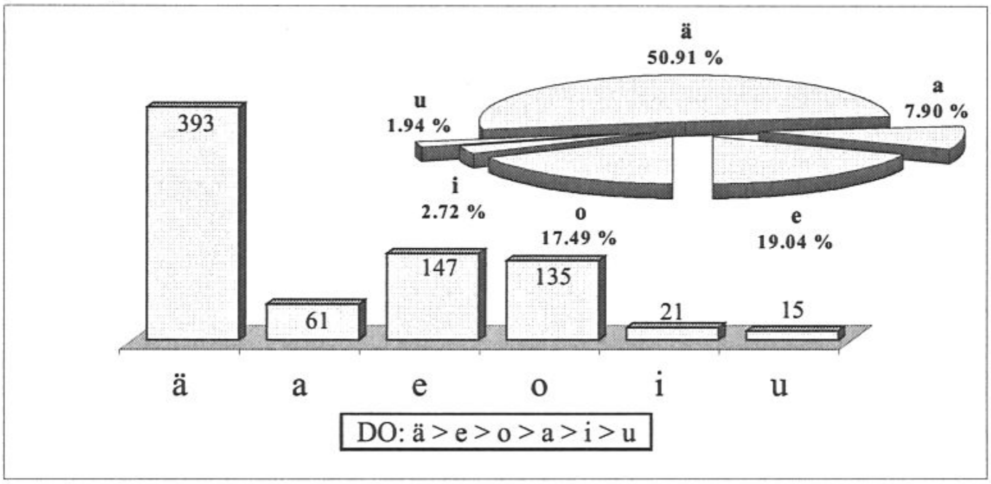
Figure 5.9: *Frequencies and percentages of place classes in the* C_2 *position of the CVCC syllables of the DBS*



5.3. Nucleus

Figure 5.10 presents vowel frequencies and percentages in the nucleus of CVCC syllables. All vowels can occur in the nucleus, but their frequencies differ greatly.

Figure 5.10: *Frequencies and percentages of vowels in the nucleus of the CVCC syllables of the DBS*

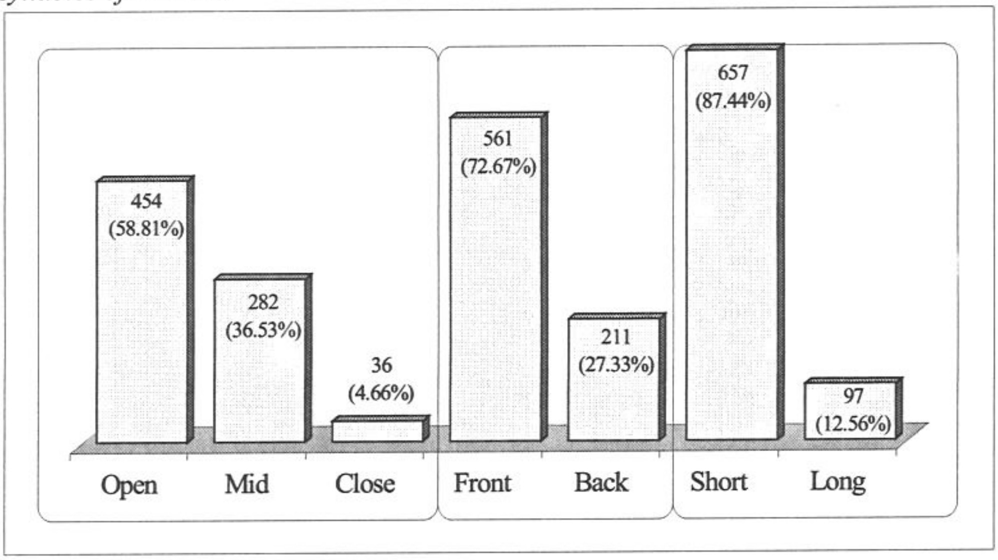


We see in the Figure 5.10 that the most frequent vowel, /ä/, covers half of all the vowel occurrences, while the proportion of the least frequent vowel, /u/, is only 2%. The theoretically maximal frequency for each vowel is 12167 (= 23x23x23). Thus, /ä/ occurs in about 3% of the maximum, for mid vowels the figure is a little over 1%, and for /u/ less than a tenth of a percent.

Mid vowels and close vowels form two clearly separate groups in terms of frequency, but it would be problematic to treat the open vowels /ä/ and /a/ as members of the same frequency group, since /ä/ is over six times as frequent as /a/. In terms of frequency, /a/ is somewhere between mid vowels and close vowels. Vowels in CVCC syllables seem to appear in two clearly separate series: short vowels (/ä, e, o/) and long vowels (/a, i, u/). In both series, there is a positive correlation between vowel frequency and sonority. As Figure 5.11 shows, a great majority (87%) of the vowels in CVCC syllables are short, and only 13% are long. The figure also shows a clear preference of front vowels.

This, however, is mainly due to the high frequency of /ä/. Finally, the figure tells us that close vowels comprise only 5% of CVCC vowels, while open vowels cover 59%.

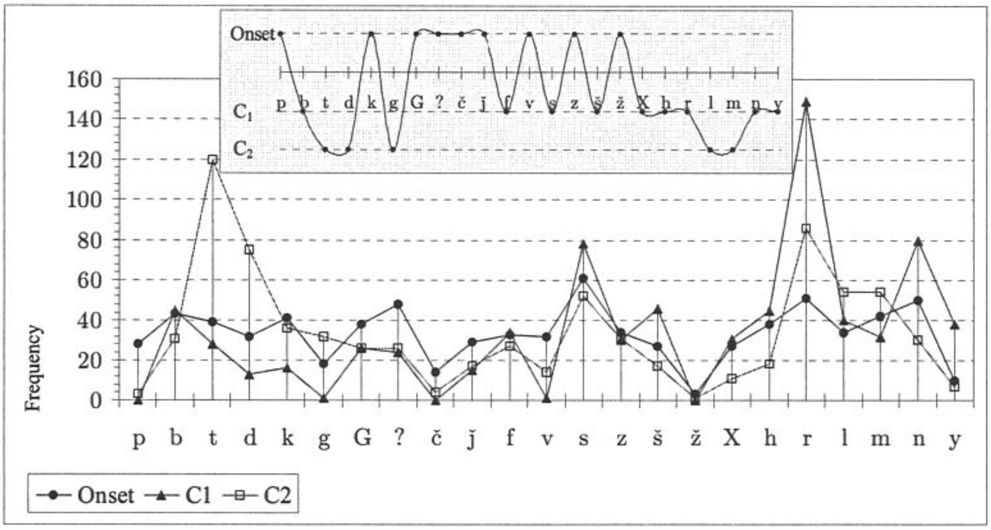
Figure 5.11: *Frequencies and percentages of vowels in the nucleus of the CVCC syllables of the DBS*



5.4. Comparison of onset and coda

Figure 5.12 compares consonant frequencies in the onset and the two coda positions. It shows that non-continuants tend to be more frequent on the edges of the syllable, i.e. in the onset and C_2. This tendency is particularly clear in the dental stops /t/ and /d/. Continuants, on the other hand, tend to be more frequent in the C_1 position, with some clear exceptions: /v/ prefers the onset, and /l/ and /m/ the C_2 position.

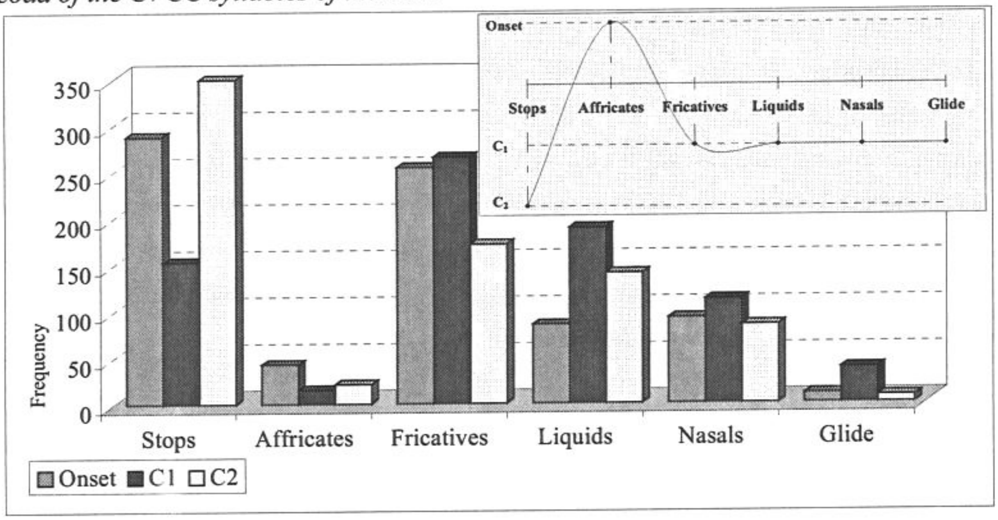
Figure 5.12: *Comparison of the frequencies of the consonants in the onset and coda of the CVCC syllables of the DBS*[43]



Differences between non-continuants and continuants can be seen more clearly in Figure 5.13, which presents the frequencies of manner subclasses in different positions of the CVCC syllable. The figure shows that stops and affricates favour the edge positions, while continuants favour the C_1 position. However, in the case of fricatives, the onset frequency is not much lower than the C_1 frequency. As was mentioned earlier, this distribution of frequencies is in harmony with the predictions of the sonority hierarchy, whereby less sonorous segments favour the edges, more sonorous segments favour the central positions.

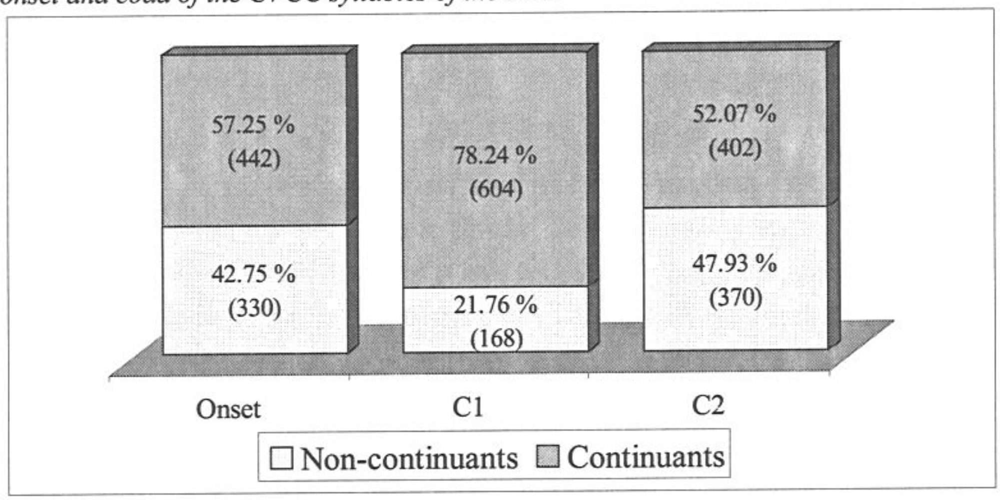
[43] The inset here and hereafter shows only the cross-positional tendencies of the manner subclasses without reference to the degree and intensity of the tendencies.

Figure 5.13: *Comparison of the frequencies of the manner subclasses in the onset and coda of the CVCC syllables of the DBS*



The percentages of non-continuants vs. continuants in different positions are given in Figure 5.14. We see that the proportion of non-continuants is particularly small in C_1, about 22%, while it is close to a half in C_2, and above 40% in the onset.

Figure 5.14: *Comparison of the frequencies of non-continuants and continuants in the onset and coda of the CVCC syllables of the DBS*



The effect of sonority can also be seen in the proportions of voiceless vs. voiced non-continuants. Since voiced segments are more sonorous than the corresponding voiceless ones, we would expect the proportion of voiced segments to increase in C_1 compared to the edge positions. This is exactly the case in non-continuants, as Figure 5.15 shows. The difference in frequencies (and percentages) between the classes of voiceless vs. voiced non-continuants are not due to the size of the class, since both classes have five consonants. However, we see that voicing in non-continuants produces a smaller difference between the C_1 and edge positions than would be the case with continuants.

Figure 5.15: *Comparison of the frequencies of voiceless and voiced non-continuants in the onset and coda of the CVCC syllables of the DBS*

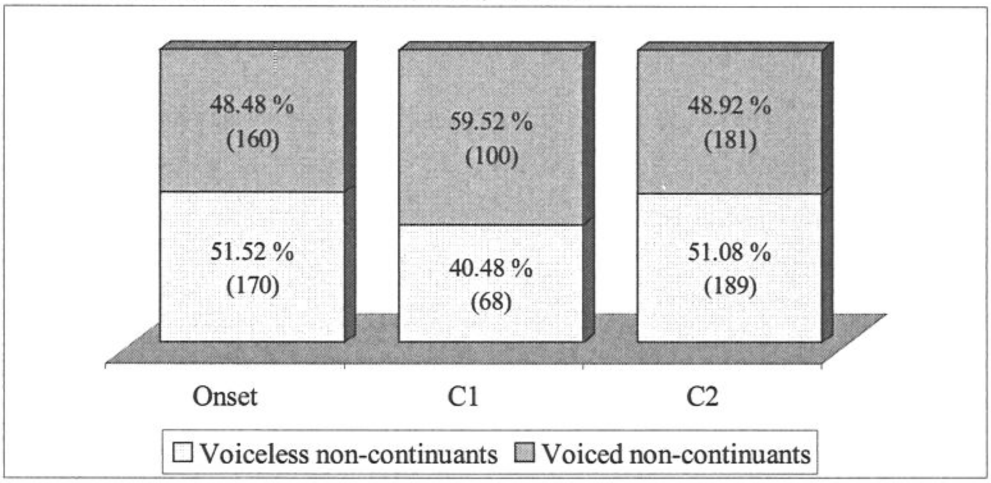
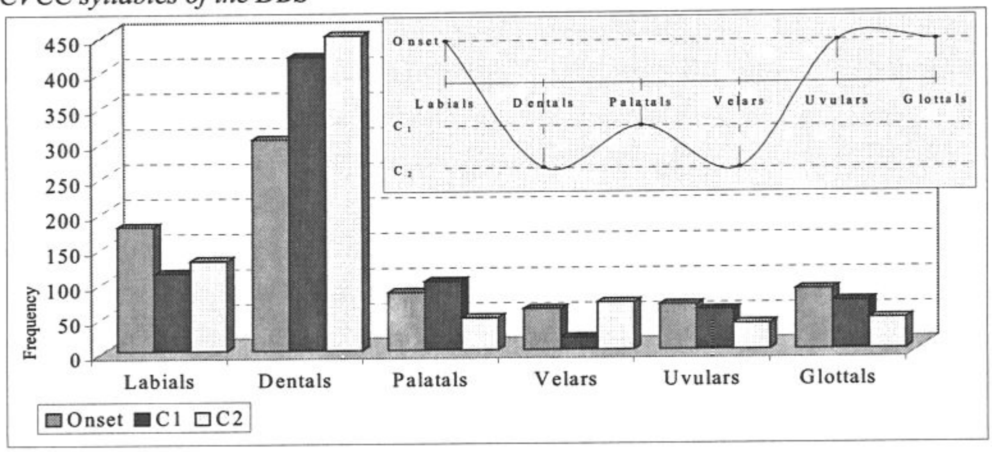


Figure 5.16 presents frequencies of place classes in different positions of the CVCC syllable. We see that dentals are in the majority, their overall frequency being more than twice the frequency of labials, which is in second biggest class. It seems that clusters in particular, i.e. in C_1 and C_2 positions, favour dentals.

Figure 5.16: *Comparison of the frequencies of place classes in the onset and coda of the CVCC syllables of the DBS*



5.5. Analysis of segment sequences in CVCC

5.5.1. Onset+nucleus

Table 5.1 presents both the occurring onset C (marked hereafter with Co)+V sequences, and gaps. The occurring sequences are marked with pluses in the table, while gaps are marked by shading. Out of the 138 (=23x6) possible Co+V sequences, 30 are missing. The table shows that five consonants occur with all vowels, i.e. three stops, /p/, /b/, and /d/, one fricative, /s/, and the lateral /l/. Most consonants have one to two gaps, usually when the nucleus is a close vowel. The largest number of gaps, four, belongs to /ž/, which only occurs when followed by an open vowel. Two consonants, /ǰ/ and /h/, occur only with short vowels.

Table 5.1: *The Co+V sequences in the CVCC syllables of the DBS (the line on the bottom and the last column to the right give the number of existing sequences)*

	p	b	t	d	k	g	G	ʔ	č	ǰ	f	v	s	z	š	ž	χ	h	r	l	m	n	y	
ä	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23
a	+	+	+	+	+	+	+	+	+		+	+	+		+	+	+		+	+	+	+	+	20
e	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	22
o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	22
i	+	+		+				+	+		+	+	+	+	+				+	+		+		13
u	+	+		+	+	+							+				+			+				8
	6	6	4	6	5	5	4	5	5	3	5	5	6	4	5	2	5	3	5	6	4	5	4	108

If we look at Table 5.1 from the point of view of the nucleus, we notice that vowels can be divided into two groups, with open and mid vowels in one group and close vowels in the other. Vowels in the first group have from zero to three gaps each. One open vowel, /ä/, can combine with all consonants. The mid vowels, have one gap each, and they do not occur with /ž/ in the onset. The other open vowel, /a/ has three gaps. In the group of close vowels, the number of gaps increases: /i/ has 10 gaps, /u/ has 15 gaps. Table 5.2 presents the numbers and percentages of existing vs. missing Co+V sequences for each vowel. It shows that over 93% of possible sequences actually occur when the nucleus is an open or a mid vowel, while less than half of the possible sequences (47%) are found when the nucleus is a close vowel. Thus, there is a positive correlation between the sonority of the vowel and the number of onset+nucleus sequences that the vowel enters into.

Table 5.2: *Number of Co+V sequences in the CVCC syllables of the DBS*

Vowel classes occupying the nucleus position	Existing combinations		Missing combinations		Total	
Open (high-sonorous) vowels	43	93.48%	3	6.52%	46	100.00%
Close-mid (mid-sonorous)	44	95.65%	2	4.35%	46	100.00%
Close (low-sonorous) vowels	21	46.65%	25	54.35%	46	100.00%
Total	108	78.26%	30	21.74%	138	100.00%

5.5.2. Nucleus+coda

Table 5.3 shows occurring and missing sequences of the nuclear vowel+C_1. Out of the 138 possible V+C_1 sequences, 56 are missing. Three consonants, /p/, /č/ and /ž/, do not occur in the C_1 position with any vowel, while /b/, /s/, /χ/ and /n/ are found in C_1 with all vowels. Some consonants, i.e. /ɢ/, /ʔ/, /ǰ/, /z/ and /h/, occur only with short vowels . Liquids and the nasal /m/ have gaps only when the nucleus is a close vowel.

Short vowels enter into more V+C_1 sequences than long vowels. When the nucleus is a short vowel, the number of occurring sequences for each vowel is 18 or 19 out of 23. But when the nucleus is a long vowel, the number of occurring sequences ranges from 12 (for /a/) to as few as 6 (for /i/).

Table 5.3: *V+C_1 combinations in the CVCC syllables of the DBS*

	p	b	t	d	k	g	G	?	č	ǰ	f	v	s	z	š	ž	χ	h	r	l	m	n	y	
ä		+	+	+	+		+	+		+	+		+	+	+		+	+	+	+	+	+	+	18
a		+	+		+						+		+		+		+		+	+	+	+	+	12
e		+	+	+	+	+	+	+		+	+		+	+	+		+	+	+	+	+	+	+	19
o		+	+	+	+		+	+		+	+	+	+	+	+		+	+	+	+	+	+		18
i		+	+										+				+			+		+		6
u		+		+	+						+		+		+		+		+			+		9
	0	6	5	4	5	1	3	3	0	3	5	1	6	3	5	0	6	3	5	5	4	6	3	82

Table 5.4 presents the occurring and missing V+C_2 combinations. In this chart of 138 slots, there are 51 gaps. All consonants can occur with at least some vowel, and four can occur with all vowels, namely, the stops /t/, /k/, and both liquids, /l/ and /r/. The following nine consonants occur in the C_2 position only if the nucleus is occupied by short vowels: /b/, /G/, /ʔ/, /ǰ/, /f/, /v/, /χ/, /h/, and /n/.

Short vowels form a fairly uniform group with regard to the number of occurring combinations which range from 22 (with /ä/ in the nucleus) to 19 (with /e/ in the nucleus). Long vowels behave less uniformly; while there are 13 combinations of V+C_2 with /a/ in the nucleus, there are as few as 6 combinations when either of the close vowels is in the nucleus.

Table 5.4: *V+C_2 combinations in CVCC syllables of the DBS*

	p	b	t	d	k	g	G	?	č	ǰ	f	v	s	z	š	ž	χ	h	r	l	m	n	y	
ä	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	22
a	+		+	+	+	+			+				+	+	+	+			+	+	+			13
e		+	+	+	+	+	+	+		+	+	+	+	+	+		+	+	+	+	+	+		19
o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+		21
i			+		+	+													+	+	+			6
u			+	+	+								+						+	+				6
	3	3	6	5	6	5	3	3	3	3	3	3	5	4	4	1	3	3	6	6	5	3	1	87

Table 5.5 presents the frequencies and percentages of existing combinations of nuclear vowels and coda consonants of CVCC syllables, as well

as gaps. As was mentioned before, open vowels are treated separately due to their differences in frequency in these combinations. We see in the upper half of the table that when the nucleus is a short vowel, around 80% of all possible V+C_1 sequences occur in the data. But with long vowels in the nucleus there is more variation; if /a/ is in the nucleus, a little over half of the possible combinations are found, while barely a third of possible sequences occur when the nucleus is a close vowel. The lower half of Table 5.5 presents the corresponding figures for V+C_2 combinations. Here, short and long vowels differ even more, i.e. with short vowels in the nucleus, the percentage of existing V+C_2 combinations ranges from 87% (for mid vowels) to 96% (for /ä/). But when a long vowel is in the nuclear position, the corresponding percentages are 57% for /a/, and only 26% for close vowels.

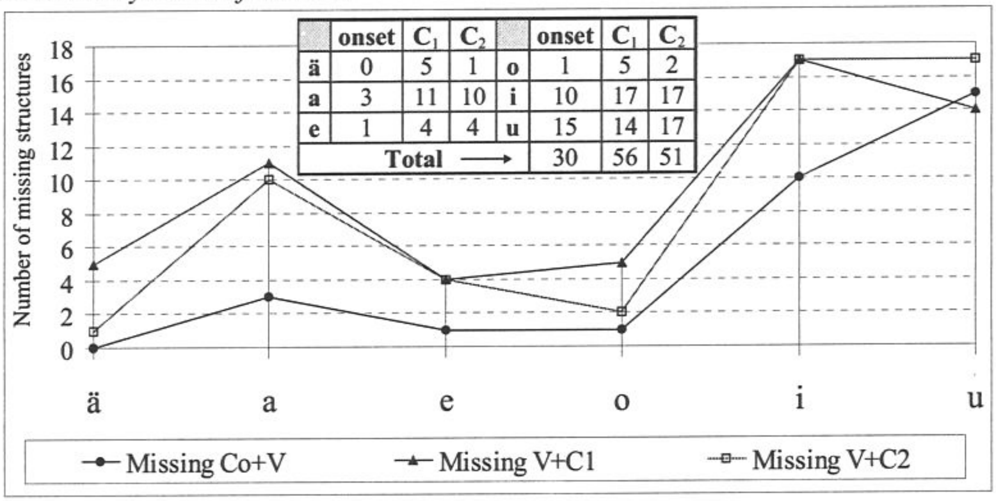
Table 5.5: *Frequencies and percentages of existing and missing V+C_1 and V+C_2 combinations in the CVCC syllables of the DBS*

A: V+C_1 combinations							
Vowel classes occupying the nucleus position		**Existing combinations**		**Missing combinations**		**Total**	
Open vowels	**ä**	18	78.26 %	5	21.74 %	23	100.00%
	a	12	52.17 %	11	47.83 %	23	100.00%
	Total	30	65.22 %	16	34.78 %	46	100.00%
Mid vowels		37	80.43%	9	19.57%	46	100.00%
Close vowels		15	32.61%	31	67.39%	46	100.00%
Total of V+C_1 sequences		82	59.42%	56	40.58%	138	100.00%
B: V+C_2 combinations							
Vowel classes occupying the nucleus position		**Existing combinations**		**Missing combinations**		**Total**	
Open vowels	**ä**	22	95.65 %	1	4.35 %	23	100.00%
	a	13	56.52 %	10	43.48 %	23	100.00%
	Total	35	76.09 %	11	23.91 %	46	100.00%
Mid vowels		40	86.96%	6	13.04%	46	100.00%
Close vowels		12	26.09%	34	73.91%	46	100.00%
Total of V+C_2 combinations		87	63.04%	51	36.96%	138	100.00%

Finally, we can look at Figure 5.17, which compares frequencies of nuclear vowels with respect to gaps in the three combinations discussed above:

Co+V, V+C_1 and V+C_2. The figure shows that for all vowels except /u/, the number of gaps is smallest in the combination of onset and nucleus. The highest number of gaps tends to be in the combination V+C_1; /u/ is also an exception here. The figure also illustrates what we saw in previous tables: with close vowels in the nucleus, the number of gaps in each combination is higher than with any other nuclear vowel.

Figure 5.17: *Comparison of the missing Co+V, V+C_1 sequences and V+C_2 structures in the CVCC syllables of the DBS*



	onset	C_1	C_2		onset	C_1	C_2
ä	0	5	1	**o**	1	5	2
a	3	11	10	**i**	10	17	17
e	1	4	4	**u**	15	14	17
Total ⟶					30	56	51

5.5.3. CC clusters in the coda

Since there are 772 syllables of the type CVCC in the DBS, the number of CC cluster tokens that they contain is thus also 772. Among these cluster tokens, there are 209 different types. The maximal number of different CC cluster types in Persian is 529 (= 23x23).[44] Thus, less than half (about 40%) of the

44 In addition to the syllable internal CC clusters, two other types of clusters are produced on syllable boundaries: medial CC clusters and medial CCC clusters. Medial CC (C-C) clusters are sequences of two consonants formed across the boundaries of syllable sequences of the following types: CVC-CVC, CVC-CV, CVC-CVCC. There are altogether 380 different medial CC clusters in the data (see Appendix 10). 202 of syllable internal CC clusters have identical counterparts among the medial CC clusters. There are 179 clusters exclusive to syllable boundaries. Five clusters, /fn/, /fχ/, /gl/, /yš/ and /zf/, which occur syllable internally, do not occur across the boundaries of syllable sequences.

theoretically possible different coda clusters are found in the data. A list of the clusters with their frequencies in the DBS is given in Appendix 9.

Table 5.6 (next page) presents all the different cluster types that were found in the data. The mark '+' in the slot means an existing cluster, whereas an empty slot indicates a missing cluster. The numbers at the edges show how many clusters each consonant is a member of, either in C_1 (the vertical column at the right edge), or in C_2 (the horizontal line at the bottom). The table shows that three consonants, /p/, /č/ and /ž/, do not occur in C_1 with any consonant, and they occur only once in C_2. Moreover, /g/ and /v/ enter into one cluster type only when in C_1 position. At the other end of the scale, /r/ is a member of 19 different clusters when in C_1, and of 16 clusters when in C_2.

Let us first look at some statistics concerning individual consonants and consonant classes that occupy each coda position, and then we can deal with properties of the clusters.

In order to compare the frequencies of each consonant in C_1 and C_2, the frequencies at the edges of Table 5.6 are presented in Figure 5.18. The shaded column indicates that the consonant occurs more often in C_1, and the white column means that the consonant is more frequent in C_2. The figure shows that stops tend to be more frequent in C_2; exceptions are two voiced stops, /b/ and /G/. Fricatives, on the other hand, tend to be a little more frequent in the C_1 position.

The absence of these five clusters from the inventory of the medial clusters seems to be accidental. As Appendix 9 shows, 21 of the medial CC clusters are sequences of identical consonants. In other words, consonantal gemination is permissible across syllable boundaries. Sequences of /ž-ž/ and /g-g/ are not found among the medial CC clusters. Medial CCC (CC-C) clusters are created in the following sequences: CVCC-CV, CVCC-CVC and CVCC-CVCC. Appendix 10 shows a list of 101 medial CCC clusters observed in the DW.

Table 5.6: *CC clusters in the CVCC syllables of the DBS*

C_1↓ C_2→	p	b	t	d	k	g	G	ʔ	č	ǰ	f	v	s	z	š	ž	χ	h	r	l	m	n	y	
p																								0
b			+	+	+		+	+					+	+	+		+	+	+	+		+		13
t		+			+		+	+			+							+	+	+	+	+	+	11
d							+					+	+					+	+	+	+			7
k			+										+						+	+	+	+		6
g																				+				1
G		+	+	+				+			+	+	+	+	+			+	+	+	+		+	14
ʔ		+	+	+							+		+	+	+				+	+	+	+	+	12
č																								0
ǰ		+		+				+				+		+				+	+		+			8
f			+				+	+				+	+	+	+		+		+	+		+	+	12
v																			+					1
s		+	+	+	+		+	+		+	+						+	+	+	+	+	+		14
z		+		+			+	+		+	+	+							+	+	+	+		11
š			+	+	+	+	+				+	+							+		+	+	+	11
ž																								0
χ			+										+	+	+				+	+	+			7
h			+	+								+	+	+	+				+	+	+	+	+	11
r		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+	+		19
l		+	+	+	+		+	+			+	+	+				+	+			+			12
m	+	+	+	+			+	+					+	+	+				+	+		+		12
n		+	+	+	+	+		+		+	+		+	+			+	+						12
y		+	+	+	+			+			+		+	+	+		+	+	+	+	+	+		15
	1	11	14	13	8	3	10	12	1	4	10	9	13	11	9	1	7	10	16	14	14	12	6	209

A conspicuous exception here is /v/, which occurs almost exclusively in C_2.[45] Sonorants, which are quite frequent in both positions, have varying preferences: /r/ and /y/ especially prefer C_1, while /l/ and /m/ are a little more

45 Samareh (1985: 169) lists fifteen clusters with /v/ in C_1 position, of which only one cluster, /vr/, is found in this study. The clusters Samareh cites – accompanied by examples in parenthesis – are as follows: /vb/ (/zovb/), /vt/ (/sovt/), /vd/ (ʔovd), /vG/ ((/zovG/) (in Samareh's transcription "zovq")) /vʔ/ (/novʔ/), /vf/ (/xovf/), /vs/ ((Govl/) (in Samareh's transcription "qovl")), /vz/ (/hovz/), /vš/ (/hoš/), /vh/ (/lovh/), /vǰ/ (/ʔovǰ/), /vm/ ((/Govm/) (in Samareh's transcription "qovm")), /vn/ (/kovn/), /vl/ (/hovl/), and /vr/ (/dovr/). The present study has interpreted the examples cited by Samareh as CVC structures and has transcribed them all without /v/. Very often in spoken Persian, the vowel /o/ is diphthongised and this diphthongisation is phonetically represented as [ow]. In my opinion, what Samareh has interpreted as /v/ is in fact the second part of this phonetic diphthong; i.e. the [w].

frequent in C_2. For some consonants, frequencies in C_1 and C_2 differ considerably, e.g. /d/, /ǰ/, /v/, and /y/, whereas differences between C_1 and C_2 frequencies are quite small for most consonants.[46]

Figure 5.18: *Comparison of consonant frequencies in the 209 different syllable internal CC clusters of the CVCC syllables of the DBS*

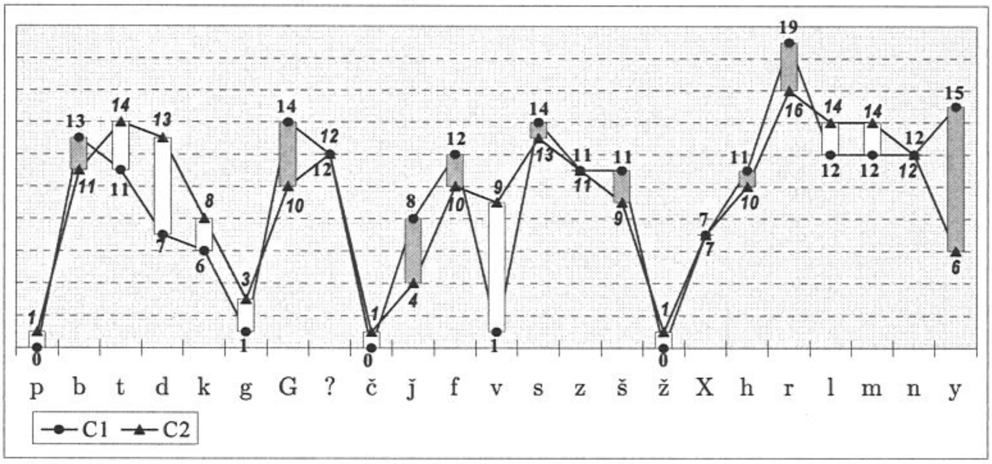


Figure 5.19 gives consonant frequencies of different manner subclasses in C_1 and C_2. The figure shows that stops and fricatives are the most frequent classes. However, if we count average frequencies per consonant in each class, sonorants turn out to be more frequent. Since there are eight stops and eight fricatives, the average frequencies per consonant in these groups range around eight and nine. On the other hand, there are only two liquids and two nasals. The average for liquids in both positions is fifteen, and for nasals twelve (in C_1) and thirteen (in C_2). The only sonorant that is below the average of stops and fricatives is the glide, when it is in C_2. According to Figure 5.19, fricatives are a

46 We can compare Figure 5.18 with Figure 5.12. Both show consonant frequencies in C_1 and C_2 of CVCC syllables. But the difference is that Figure 5.12 is based on all 772 cluster tokens, whereas Figure 5.18 is based on the 209 different cluster types. By comparing the figures, we can see some effects of cluster frequencies. In Figure 5.18, /G/ and /ʔ/ are quite frequent compared to other consonants, but they are not among the peaks in Figure 5.12 We can interpret this to mean that while /G/ and /ʔ/ enter into many different cluster types, they tend to occur in lower frequency clusters. On the other hand, /t/, /s/ and /r/ are frequent in both figures. This means that they occur both in numerous cluster types and in higher frequency clusters.

little more frequent in C_2. However, if we look back at Figure 5.18, we see that the difference is mainly due to the deviant behaviour of /v/ (and also of /z/), since all other fricatives are either more frequent in C_1, or have identical frequencies in both positions.

Figure 5.19: *Comparison of the frequencies of manner subclasses in C_1 and C_2 positions of the 209 different syllable internal CC clusters in the CVCC syllables of the DBS*

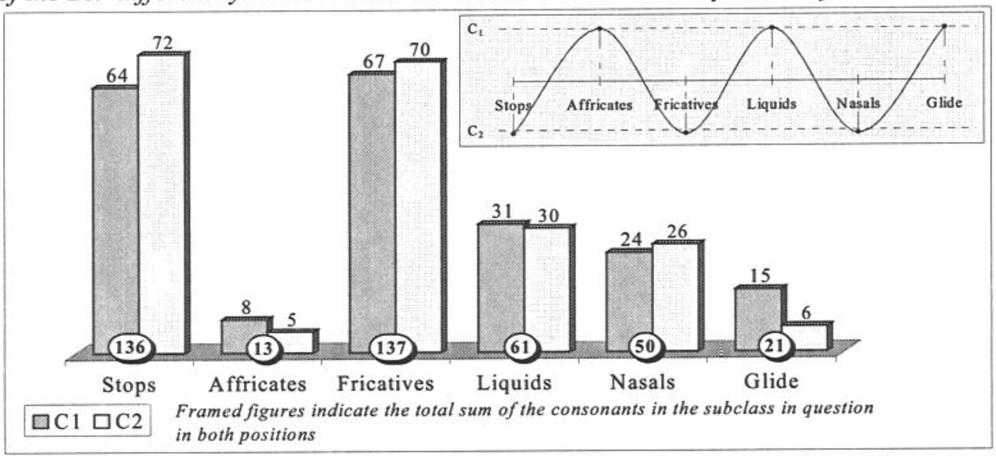
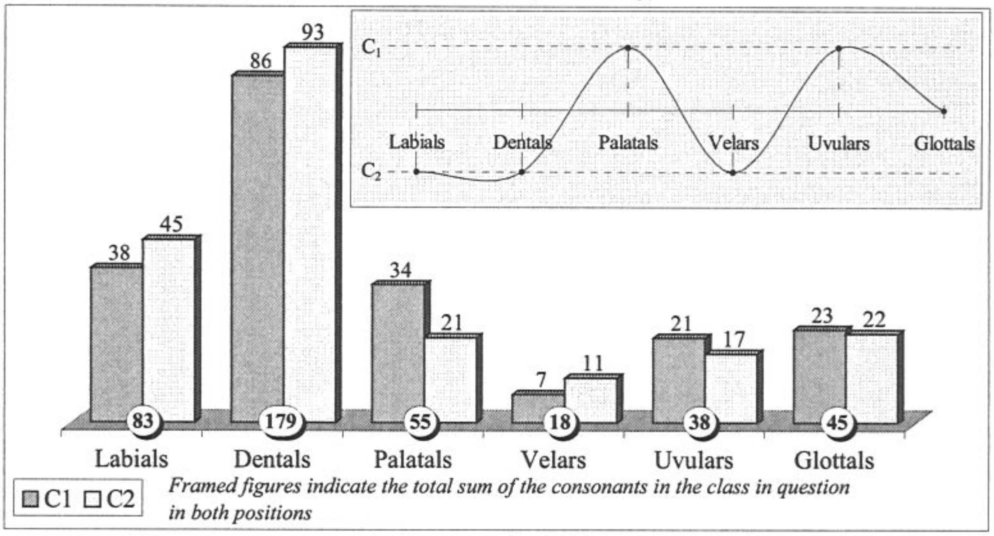


Figure 5.20 shows frequencies of each place class. It shows that dentals are the most frequent group, with labials second. Looking back at Figure 5.18, we see that dentals in each manner class tend to have fairly high frequencies, whereas frequencies of labial consonants vary greatly. The least frequent group is velars. Part of the explanation for the low velar frequencies is that there are only two velars. However, uvulars and glottals are clearly more frequent than velars in spite of the fact that each class has two members, as velars do. But both of the velars are stops, which do not favour C_1, whereas uvulars and glottals each have one fricative in addition to a stop. In other words, place frequencies also reflect the frequencies of manner classes.

Figure 5.20: *Comparison of the frequencies of place classes in C_1 and C_2 positions of the syllable internal CC clusters in the CVCC syllables of the DBS*



Let us next have a closer look at the nature of CC clusters themselves. The sonority hypothesis predicts that, sonority decreases towards the syllable edges within a syllable, i.e. the consonant in C_2 is less sonorous than the preceding consonant in C_1. Since obstruents are less sonorous than resonants, we would expect that the combination *sonorant+obstruent* would be the preferred one, whereas the opposite sequence, *obstruent+sonorant,* would be disfavoured. Those sequences where the sonority remains the same, i.e. *obstruent+obstruent* and *sonorant+sonorant,* would be somewhere in between. We can compare the sequences in terms of their preference by counting the percentage of existing sequences out of theoretically possible sequences. Table 5.7 shows the four different sequences with their frequencies and the percentage of each sequence out of the theoretically possible ones. As expected, the sequence *sonorant+obstruent* has the highest percentage. What is unexpected, however, is that the second highest percentage belongs to the reversed sequence, *obstruent+sonorant,* where the sonority increases rather than decreases towards the syllable edge. In those sequences, where the sonority (measured by the *obstruent/sonorant* distinction) remains the same, the percentage of theoretically possible sequences is lowest.

Table 5.7: *Different combinations of the manner classes in the CC clusters in the CVCC syllables of the DBS*

Combinations	Existing clusters		Missing clusters		Total	
obstruent+obstruent	87	26.85%	237	73.15%	324	100.00%
sonorant+sonorant	10	40.00%	15	60.00%	25	100.00%
obstruent+sonorant	52	57.78%	38	42.22%	90	100.00%
sonorant+obstruent	60	66.67%	30	33.33%	90	100.00%
Total	209	39.51%	320	60.49%	529	100.00%

One might think that the odd picture suggested by Table 5.7 results from an erroneous treatment of how sonority works in Modern Persian. We saw above that the distributions in C_1 and C_2 of fricatives, i.e. continuant obstruents, are closer to the distributions of sonorants rather than to those of non-continuant obstruents (stops and affricates). Therefore, one could think that the categories non-continuants (stops and affricates) vs. continuants (fricatives and sonorants) would give a more reliable picture of the sequences with respect to sonority. This, however, makes the picture even more difficult to interpret; it turns out that the most favoured sequence, the one with the highest proportion (48%) of existing sequences out of the theoretically possible ones belongs to the sequence type *continuant+continuant,* where sonority (measured by the categories continuant/non-continuant) remains the same.

In order to see what is going on here, consonants are divided into three groups, non-continuant obstruents, continuant obstruents and sonorants, and the sequences they form are studied. The frequencies of different sequence types are calculated, and for each sequence type the percentage of existing sequences out of the theoretically possible maximum is presented. The results are given in Table 5.8. The clusters are presented in three groups: lines 1 to 5 contain sequences with non-continuant obstruents as a member, lines 6 to 10 show sequences with continuant obstruents as one member, and lines 11 to 15 have sequences with sonorants as one member. For the ease of comparison, some sequence types are repeated in different groups (i.e. lines 2 and 3 = lines 7 and 8; lines 4 and 5 = lines 12 and 13; and lines 9 and 10 = lines 14 and 15).

In the first group, line 1 gives figures of consonant sequences where both consonants are non-continuant obstruents. We see that the percentage of existing clusters is lower here than in any other cluster type in the first group, or in any other group for that matter. The percentage of existing clusters increases when the other member of the cluster is a continuant obstruent (lines 2 and 3), and it increases even more when the other member is a sonorant (lines 4 and 5). Moreover, the percentages are higher in clusters where sonority decreases towards the syllable edge (line 3 vs. line 2, and line 5 vs. line 4).

Table 5.8: *Combinations of manner subclasses in 209 different CC clusters of CVCC syllables in the DBS*[47]

Clusters →		**Existing**		**Missing**		**Total**	
No.	**Combinations**	**freq.**	**%**	**freq.**	**%**	**freq.**	**%**
1	**non-cont. obst.+non-cont. obst.**	21	21	79	79	100	100
2	**non-cont. obst.+cont. obst.**	21	33	43	67	64	100
3	**cont. obst.+non-cont. obst.**	23	36	41	64	64	100
4	**non-cont. obst.+son.**	27	54	23	46	50	100
5	**son.+non-cont. obst.**	33	66	17	34	50	100
6	**cont. obst.+cont. obst.**	19	30	45	70	64	100
7	**non-cont. obst.+cont. obst.**	21	33	43	67	64	100
8	**cont. obst.+non-cont. obst.**	23	36	41	64	64	100
9	**cont. obst.+son.**	25	62	15	38	40	100
10	**son.+cont. obst.**	27	68	13	32	40	100
11	**son.+son.**	10	40	15	60	25	100
12	**non-cont. obst.+son.**	27	54	23	46	50	100
13	**son.+non-cont. obst.**	33	66	17	34	50	100
14	**cont. obst.+son.**	25	62	15	38	40	100
15	**son.+cont. obst.**	27	68	13	32	40	100

Lines 6 to 10, i.e. combinations with continuant obstruents as one member, confirm what was just said: the percentage of existing clusters is lowest when both members are continuant obstruents (line 6), and the percentages are highest when a sonorant is the other member (lines 9 and 10). Moreover, the percentages are higher for clusters where sonority decreases towards the syllable edge (i.e. line 8 (= line 3) has a higher percentage than line 7 (= line 2), and line 10 also has a higher percentage than line 9).

The last group (lines 11 to 15) shows clusters with a sonorant consonant as one of the members. We see the same tendencies as above. First, combinations of

[47] cont. = continuant, non-cont. = non-continuant, obs. = obstruent, son.= sonorant, freq.= frequency.

dissimilar consonants have higher percentages than clusters consisting of similar consonants (i.e. the percentage of line 11 is the lowest in the third group). Second, the percentages increase in combinations where sonority decreases towards the syllable edge. In other words, there seem to be two factors at work here, one is a tendency towards *dissimilarity* of consonants in CC clusters, and the other is a tendency towards *decreasing sonority* at the syllable edge. What counts as dissimilar is most clearly obstruents vs. sonorants, but to some extent also non-continuant vs. continuant obstruents (compare line 1 to lines 2 and 3). If the combinations of maximally dissimilar consonants were the preferred ones, we would expect the clusters of *sonorant+non-continuant obstruent* to have the highest percentage. However, the highest percentage out of the theoretically possible combinations belongs to the cluster type *sonorant+continuant obstruent,* but we have to note that the difference between the percentages is really slight and probably not significant (68% for *sonorant+continuant obstruent* and 66% for *sonorant+non-continuant obstruent).*

The tendency towards dissimilarity is also illustrated in the data of Table 5.9. The table shows sequences of consonants from the same manner subclass, with frequencies and percentages of existing vs. missing clusters out of the theoretically possible ones. The table shows that affricates, liquids and the glide do not cluster at all with consonants from the same group. Moreover, those that form clusters with their likes, i.e. stops, fricatives and nasals, do so in less than 30% of the cases.

Table 5.9: *CC clusters with members belonging to identical manner subclasses (based on Appendix 12)*

Clusters	**Existing clusters**		**Missing clusters**		**Total**	
stop+stop	18	28.13%	46	71.88%	64	100.00%
affricate+affricate	0	0.00%	4	100.00%	4	100.00%
fricative+fricative	19	29.69%	45	70.31%	64	100.00%
liquid+liquid	0	0.00%	4	100.00%	4	100.00%
nasal+nasal	1	25.00%	3	75.00%	4	100.00%
glide+glide	0	0.00%	1	100.00%	1	100.00%

The tendency towards dissimilarity is not restricted to manner classes. Table 5.10 shows clusters whose members come from the same place class. We see that, with the exception of dentals, consonants of the same place class do not easily form clusters. Velars, uvulars and glottals do not combine at all with consonants of the same place, and labials and palatals form such clusters in about 10% of the possible cases only. Dentals, on the other hand, combine with dentals in a little more than half of the theoretically possible cases.

Table 5.10: *CC structures with members belonging to identical place classes (based on Appendix 13)*

Subgroups		**Existing clusters**		**Missing clusters**		**Total**	
labials	**labials**	3	12.00%	22	88.00%	25	100.00%
dentals	**dentals**	27	55.10%	22	44.90%	49	100.00%
palatals	**palatals**	2	8.00%	23	92.00%	25	100.00%
velars	**velars**	0	0.00%	4	100.00%	4	100.00%
uvulars	**uvulars**	0	0.00%	4	100.00%	4	100.00%
glottals	**glottals**	0	0.00%	4	100.00%	4	100.00%

Before going into individual clusters in detail, let us have a look at the role of voicing in obstruent clusters. If voicing differences were an important factor in the dissimilarity of cluster members we would expect those clusters whose members differ in voicing to be the preferred ones. And if voicing (in obstruents) were an important sonority factor, we would expect the sequence type *voiced+voiceless* to be the most frequent one. Table 5.11 presents the frequencies of different manner subclass sequences. Those clusters whose frequencies are zero, are left out (i.e. the sequences *stop+affricate* and *affricate+affricate* are missing). The frequencies of each cluster type are divided into four groups on the basis of the voicing of the cluster consonants. The frequencies in each slot are quite small, but big enough to show something of the role of voicing. We see, first, that the preferred voicing sequence varies in different manner subclass clusters. Thus, the sequence *voiced+voiceless* is the most common type of *stop+fricative* clusters, while the sequence of two voiceless consonants is the preferred one in the combination of two fricatives. Furthermore, the sequence *fricative+stop* seems to have two favoured voicing combinations, *voiceless+voiceless* and *voiceless+voiced,* while the sequence *stop+stop* does not have any favoured voicing combination. Second, the table shows that the voicing sequence type with the highest frequency (29) consists of clusters of two voiceless consonants; all the other voicing sequence types are roughly equally frequent (18-20). This shows that voicing (in obstruents) is not important for the sonority difference or the dissimilarity of the cluster members.[48]

Table 5.11: *Subgroups of the obstruent+obstruent CC clusters in the CVCC syllables of the DBS with regard to the voicing of their constituting segments*

Cluster types	**voiced+ voiced**	**voiceless+ voiceless**	**voiced+ voiceless**	**voiceless+ voiced**	**Total**
stop+stop	5	4	5	4	18
stop+fricative	4	6	10	1	21
affricate+stop	2	0	1	0	3
affricate+fricative	2	0	1	0	3
fricative+stop	3	9	1	8	21
fricative+affricate	1	0	0	1	2
fricative+fricative	1	10	2	6	19
Total	18	29	20	20	87

Let us now go back to the data in Table 5.6 and see in detail which consonants form clusters with each other, and what the cluster frequencies are, i.e. in how many syllables of the DBS each of the 209 clusters occur. As was mentioned before, Appendix 9 list clusters with their frequencies. The appendix shows that the frequencies range from 1 to 33; a little more than a third of the clusters (35%) occur only once. In the following discussion, clusters are divided into three groups. Clusters whose frequencies are ten or more will be called *high-frequency clusters.* There are forteen such clusters, which is 6.70% of all clusters. Frequencies that range between nine and five will be called *moderate;* there are 30 clusters (14.35%) with such frequencies. The rest, i.e. clusters whose frequencies are four or less, will be referred to as *low-frequency clusters.* The vast majority, 165 clusters (78.95%), belong to this groups. The division is no

[48] We saw earlier in Figure 5.15 some preference to voiced non-continuants over voiceless ones in C_1. The difference between the data in Figure 5.15 and Table 5.11 is that Figure 5.15 is based on the 772 cluster tokens in the DBS, while Table 5.12 is based on the 209 different clusters in the DBS. We get a preference of voiced non-continuants in C_1 as against voiceless ones in Table 5.11 as well: of the non-continuant consonants in C_1, 67% are voiced and 33% voiceless; of the non-continuant consonants in C_2, only 54% are voiced and 46% are voiceless. The point of Table 5.11 is to show that there is no systematic preference with respect to voicing on the level of combinations of different manner subclasses.

doubt arbitrary, but its only purpose is to give some idea of the frequency variation of clusters.

To make the description easier, Table 5.6 is divided into subtables. We can start with obstruent clusters. Obstruents in turn are divided into subclasses (i.e. stops, affricates, and fricatives). Different types of *obstruent+obstruent* clusters and their frequencies are given in Tables 5.12 to 5.18. It is assumed here and in comments of subsequent tables of this chapter that all manner subclasses represent different degrees of sonority, whereas all members of the same manner subclass have the same degree of sonority.

Table 5.12: *Stop+stop clusters in the CVCC syllables of the DBS*

$C_1\downarrow C_2\rightarrow$	p	b	t	d	k	g	G	ʔ	Clusters and their frequencies		
p									bt 5	ʔb 1	tʔ 1
b			+	+	+		+	+	ʔd 4	ʔt 1	tb 1
t		+			+		+	+	Gd 3	bd 1	Total of clusters 18
d							+		tG 3	dG 1	Total of tokens 33
k			+						bʔ 2	kt 1	
g									bk 2	Gb 1	
G		+	+	+				+	Gt 2	bG 1	
ʔ		+	+	+					tk 2	Gʔ 1	

Table 5.13: *Stop+fricative clusters in the CVCC syllables of the DBS*

$C_1\downarrow C_2\rightarrow$	f	v	s	z	š		χ	h	Clusters and their frequencies		
p									bh 5	ds 2	dh 1
b			+	+	+		+	+	ks 5	Gf 2	dv 1
t	+							+	bs 3	Gs 2	Gš 1
d		+	+					+	bz 3	th 2	Gh 1
k			+						Gz 3	ʔš 1	Gv 1
g									tf 3	ʔf 1	Total of clusters 21
G	+	+	+	+	+			+	ʔs 2	bš 1	Total of tokens 43
ʔ	+		+	+	+				ʔz 2	bχ 1	

Table 5.14: *Affricate+stop clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	p	b	t	d	k	g	G	ʔ	Clusters and their frequencies
č									ǰb 2 ǰʔ 1 Total of clusters 3
ǰ		+		+				+	ǰd 2 Total of tokens 5

Table 5.15: *Affricate+fricative clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	f	v	s	z	š		χ	h	Clusters and their frequencies
č									ǰh 1 ǰz 1 Total of clusters 3
ǰ		+		+				+	ǰv 1 Total of tokens 3

Table 5.16: *Fricative+stop clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	p	b	t	d	k	g	G	ʔ	Clusters and their frequencies
f			+				+	+	st 33 fʔ 3 fG 1
v									št 22 ht 3 sʔ 1
s		+	+	+	+		+	+	χt 17 zb 3 sd 1
z		+		+			+	+	ft 15 zd 3 sG 1
š			+	+	+	+	+		šk 9 šG 2 zG 1
									sb 6 zʔ 2 Total of clusters 21
χ			+						sk 5 šd 1 Total of tokens 134
h			+	+					hd 4 šg 1

Table 5.17: *Fricative+affricate clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	č	ǰ	Clusters and their frequencies
f			
v			sǰ 1 Total of clusters 2
s		+	zǰ 1 Total of tokens 2
z		+	
š			
χ			
h			

Table 5.18: *Fricative+fricative clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	f	v	s	z	š		χ	h	Clusters and their frequencies		
f		+	+	+	+		+		hv 4	χs 2	χz 1
v									fš 3	šf 1	zf 1
s	+						+	+	sχ 3	šv 1	zv 1
z	+	+							χš 3	fv 1	Total of clusters 19
š	+	+							fs 2	fχ 1	Total of tokens 33
									hš 2	fz 1	
χ			+	+	+				hs 2	hz 1	
h		+	+	+	+				sf 2	sh 1	

There are nine (3x3) theoretically possible types of obstruent subclass clusters. Two types do not occur at all, namely, *stop+affricate* and *affricate+affricate.* Moreover there are relatively few clusters with an affricate as a member, and their cluster frequencies are conspicuously low, with two at most as Tables 5.14, 5.15 and 5.17 show. The remaining four tables (Tables 5.12, 5.14, 5.17 and 5.19) are fairly similar in the number of clusters, which ranges from 18 to 21, but there are differences in cluster frequencies. The cluster type *fricative+stop* (Table 5.16) is the only one that has high frequency clusters, such as /st/ with a frequency of 33, whereas the highest frequency is only five in *stop+stop* clusters (Table 5.12) and *stop+fricative* clusters (Table 5.13), and as low as four in *fricative+fricative* clusters (Table 5.18). Consequently, the total frequencies of the cluster types differ dramatically, thus the total of tokens for *fricative+stop* clusters is 134 (Table 5.16), whereas the total frequencies for the other three groups are only a third or fourth of that figure, ranging from 43 (*stop+fricative*, Table 5.13) to 33 (*stop+stop*[49], Table 5.12, and

49 Samareh (1985: 142) has introduced 14 two-stop clusters. His data does not include three monosyllable CVCC words: /tebG/ (in his transcription /tebq/), /väGʔ/ (in his transcription /vaqʔ/) and /näʔt/ (in his transcription /naʔt/) and neither does he mention any CVCC syllable with the /kt/ cluster. Thus, four clusters, /bG/, /Gʔ/, /ʔt/ and /kt/, are missing in his list of stop+stop clusters. The first two words have been excluded, as he states, because they are always used together with a suffix. In my opinion, this reason is not strong enough to exclude the syllables /tebG/ and /väGʔ/. Both in Samareh's research and in this work the attention is primarily focused on the structure of syllables and not words. In other words, /tebG/ and /väGʔ/ should be dealt with as syllables and not words. Secondly, both of the words may also be used independently without any suffix. For example:

fricative+fricative Table 5.18). It is interesting to note that the cluster type *fricative+stop,* which seems to be the favoured one among the *obstruent+obstruent* clusters, is also a sequence where sonority decreases towards the syllable edge. Moreover, it is the only sequence type here with decreasing sonority if we disregard affricates, which are quite infrequent in clusters.

Let us next look at *sonorant+sonorant* sequences. Sonorant subgroup clusters and their frequencies are given in Tables 5.19-5.23. Out of the nine (3x3) possible subclass cluster types, four do not occur at all, namely *nasal+glide, liquid+glide, liquid+liquid* and *glide+glide.* In the remaining five cluster types, frequencies are not very high, with the highest frequency, 12, belonging to /rm/ (Table 5.19). We saw before that there is a tendency towards dissimilarity in cluster consonants, and clusters of sonorants conform this tendency. First, there is only one cluster, /mn/ (Table 5.20), where consonants come from the same manner subclass, and its frequency is as low as 3. On the other hand, combinations where consonants come from different manner subgroups are favoured: all the possible *glide+other sonorant* clusters occur (Tables 5.22-23), and three out of four possible *liquid+nasal* clusters are found (Table 5.19). These (relatively) dissimilar combinations also include clusters whose frequencies are the highest among *sonorant+sonorant* sequences.

In addition to the tendency towards dissimilarity, we also see evidence that those clusters where the sonority decreases towards the syllable edge are favoured. First, we see that the other two cluster types that are missing, namely, *nasal+glide* and *liquid+glide,* have increasing (and not decreasing) sonority towards the syllable edge. Second, there is one pair where consonants occur in

---

a. /ʔin važe ra tebG täläffoz mikonänd/. 'This word is pronounced *tebG*'.
b. /väGʔ nähadän be härfe här käso nakäsi šayeste nist/. "It is not appropriate to pay attention to whatever every noble and ignoble person says).

With regard to the exclusion of the monosyllabic word /näʔt/, Samareh (1985) states that it "is an obsolete literary word and is not used – neither in speech nor in writing – in Modern Persian". It is true that the word in question is a literary word, but I disagree with the statement that it is "obsolete" and "is not used". Moreover, my impression of Samareh's data is that his work is similar to the present study and is mainly based on literary Persian and not colloquial Persian. Samareh does not mention the reason for the exclusion of the cluster /kt/.

both orders, namely, *liquid+nasal (*Table 5.19) and *nasal+liquid* (Table 5.21), which differ especially in cluster frequencies but also in the number of cluster types. The sequence *nasal+liquid* with increasing sonority is found in two out of four possible clusters (Table 5.20), and none of them is a high frequency cluster, whereas the opposite order, *liquid+nasal* has three clusters, and one of them, /rm/, has a frequency of 12.

Table 5.19: *Liquid+nasal clusters in the CVCC syllables of the DBS*

$C_1\downarrow C_2\rightarrow$	m	n	Clusters and their frequencies
r	+	+	rm 12 rn 2 Total of clusters 3
l	+		lm 4 Total of tokens 18

Table 5.20: *Nasal+nasal clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	m	n	Clusters and their frequencies
m		+	mn 3 Total of clusters 1
n			Total of tokens 3

Table 5.21: *Nasal+liquid clusters in the CVCC syllables of the DBS*

$C_1\downarrow C_2\rightarrow$	r	l	Clusters and their frequencies
m	+	+	mr 7 Total of clusters 2
n			ml 2 Total of tokens 9

Table 5.22: *Glide+liquid clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	r	l	Clusters and their frequencies
y	+	+	yl 10 Total of clusters 2
			yr 4 Total of tokens 14

Table 5.23: *Glide+nasal clusters in the CVCC syllables of the DBS*

$C_1\downarrow C_2\rightarrow$	m	n	Clusters and their frequencies
y	+	+	yn 6 Total of clusters 2
			ym 1 Total of tokens 7

In the next section, Tables 5.24-5.31 present clusters and frequencies of the type *sonorant+obstruent.* Since cluster consonants here come from different subgroups and since sonority in these sequences decreases towards the syllable edge, we would expect this type to be the favoured one in the sense that it has both a high number of clusters and high cluster frequencies. The tables show that this is indeed the case. Out of the nine (3x3) possible combinations of subclasses, only one is missing, namely, *glide+affricate*. Of the existing combinations, the type *liquid+stop* has the highest total frequency of cluster tokens, 98 (Table 5.27). There are two high frequency clusters in this group, /rd/ with a frequency of 26, and /rt/, with a frequency of 11. Of the remaining 11 clusters, 7 have moderate frequencies ranging from 9 to 5. The next highest frequency of cluster tokens, 71, belongs to combinations of *nasal+stop* (Table 5.24). There are 12 clusters of this type, and two of them have high frequencies, i.e. /ng/ whose frequency is 26, and /nd/ whose frequency is 22. All the other clusters here have low frequencies, at most 4, and six clusters, i.e. half of the total of 12, have frequencies as low as 2. If we disregard the combinations of *glide+obstruent* (Tables 5.30, 5.31) on account of their low overall frequencies, we can say that stops are the favoured subclass as C_2 in *sonorant+obstruent* sequences.

The third highest number of cluster tokens, 63, is found in *liquid+fricative* sequences (Table 5.29). Two high-frequency clusters are /rs/, whose frequency is 17, and /rz/ with a frequency of 14. Only one cluster, /rf/, has a moderately high frequency of 9; all the other clusters of the 13 total have low frequencies, at most 4.

Next in order is the sequence *nasal+fricative* (Table 5.26), with a total of 8 clusters and 20 tokens. None of these clusters have a high frequency, and only one has a moderate frequency of 9. Half of the clusters occur only in one syllable.

It is interesting to note that sequences of *sonorant+affricate* also have higher frequencies than other combinations with affricates. Table 5.25 shows that the only *nasal+affricate* cluster, /nǰ/, has a moderately high frequency of 9, and Table 5.28 shows that one of the two clusters, /rǰ/, has a frequency of 6, the other cluster, /rč/, has a frequency of 4. All the other clusters with an affricate as a member have a frequency of at most 4, but most combinations with an affricate have a frequency of 1(see Tables 5.15 and 5.17).

Table 5.24: *Nasal+stop clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	p	b	t	d	k	g	G	ʔ	Clusters and their frequencies
m	+	+	+	+			+	+	ng 26 mʔ 4 nt 3 md 2 nʔ 2 nk 2 Total of clusters 12
n		+	+	+	+	+		+	nd 22 mp 3 mb 2 mG 2 nb 2 mt 1 Total of tokens 71

Table 5.25: *Nasal+affricate clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	č	ǰ	Clusters and their frequencies
m			nǰ 9 Total of clusters 1
n		+	Total of tokens 9

Table 5.26: *Nasal+fricative clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	f	v	s	z	š		χ	h	Clusters and their frequencies
m			+	+	+				ns 9 mz 2 mš 1 nh 1 Total of clusters 8
n	+		+	+			+	+	ms 3 nz 2 nf 1 nχ 1 Total of tokens 20

Table 5.27: *Liquid+stop clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	p	b	t	d	k	g	G	ʔ	Clusters and their frequencies
r		+	+	+	+	+	+	+	rd 26 rk 9 rG 8 lk 5 rg 5 lb 4 ld 3 Total of clusters 13
l		+	+	+	+		+	+	rt 11 rb 8 lG 6 lt 5 lʔ 4 rʔ 4 Total of tokens 98

Table 5.28: *Liquid+affricate clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	č	ǰ	Clusters and their frequencies
r	+	+	rǰ 6 Total of clusters 2
l			rč 4 Total of tokens 10

Table 5.29: *Liquid+fricative clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	f	v	s	z	š		χ	h	Clusters and their frequencies
r	+	+	+	+	+	+	+	+	rs 17 rf 9 lf 3 rv 3 lh 2 lv 1 rž 1 Total of clusters 13
l	+	+	+				+	+	rz 14 rš 4 rh 3 rχ 3 ls 2 lχ 1 Total of tokens 63

Table 5.30: *Glide+stop clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	p	b	t	d	k	g	G	ʔ	Clusters and their frequencies
y		+	+	+	+			+	yd 3 yʔ 1 yt 1 Total of clusters 5 yk 2 yb 1 Total of tokens 8

Table 5.31: *Glide+fricative clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	f	v	s	z	š		χ	h	Clusters and their frequencies
y	+		+	+	+		+	+	yf 4 yh 1 yχ 1 Total of clusters 6 yš 1 ys 1 yz 1 Total of tokens 9

Finally, let us look at clusters of the opposite order, i.e. *obstruent+sonorant,* presented in Tables 5.32-5.39. Since sonority in these sequences increases towards the syllable edge, we would expect lower cluster and token frequencies than in the previous section. As we will see, this is true, with one exception.

Table 5.32: *Stop+nasal clusters in the CVCC syllables of the DBS*

C_1↓ C_2→	**m**	**n**	**Clusters and their frequencies**		
p			ʔn 3	kn 1	Total of clusters 9
b		+	ʔm 2		Total of tokens 16
t	+	+	bn 2		
d	+		Gm 2		
k	+	+	tm 2		
g			tn 2		
G	+		dm 1		
ʔ	+	+	km 1		

Table 5.33: *Stop+liquid clusters in the CVCC syllables of the DBS*

C_1↓ C_2→	**r**	**l**	**Clusters and their frequencies**		
p			br 12	ʔr 2	Total of clusters 13
b	+	+	tr 8	kl 2	Total of tokens 58
t	+	+	bl 7	dl 1	
d	+	+	dr 6	gl 1	
k	+	+	kr 6	Gr 1	
g		+	Gl 5		
G	+	+	ʔl 4		
ʔ	+	+	tl 3		

Table 5.34: *Stop+glide clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	y	Clusters and their frequencies		
p		ʔy 1	Total of clusters	3
b		ɢy 1	Total of tokens	3
t	+	ty 1		
d				
k				
g				
ɢ	+			
ʔ	+			

Table 5.35: *Affricate+nasal clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	m	n	Clusters and their frequencies		
č			ǰm 3	Total of clusters	1
ǰ	+			Total of tokens	3

Table 5.36: *Affricate+liquid clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	r	l	Clusters and their frequencies		
č			ǰr 4	Total of clusters	1
ǰ	+			Total of tokens	4

Table 5.37: *Fricative+nasal clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	m	n	Clusters and their frequencies			
f		+	sm 9	fn 1	Total of clusters	10
v			zm 6	sn 1	Total of tokens	37
s	+	+	hn 5			
z	+	+	šm 4			
š	+	+	hm 4			
			χm 3			
χ	+		zn 3			
h	+	+	šn 1			

Table 5.38: *Fricative+liquid clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	r	l	**Clusters and their frequencies**		
f	+	+	hr 14	χl 3	Total of clusters 12
v	+		sr 8	fl 2	Total of tokens 55
s	+	+	sl 6	χr 2	
z	+	+	zr 5	vr 1	
š	+		hl 4		
			zl 4		
χ	+	+	šr 3		
h	+	+	fr 3		

Table 5.39: *Fricative+glide clusters in the CVCC syllables of the DBS*

C₁↓ C₂→	y	**Clusters and their frequencies**	
f	+	hy 2	Total of clusters 3
v		šy 1	total of tokens 4
s		fy 1	
z			
š	+		
χ			
h	+		

Out of the nine (3x3) possible manner subclass sequences, only one is not found, namely, *affricate+glide.* The sequences that have the highest number of clusters and the highest total of cluster tokens are *stop+liquid* (Table 5.33), with 13 clusters and 58 cluster tokens, and *fricative+liquid* (Table 5.38), with 12 clusters and 55 cluster tokens. Both groups contain only one cluster where the frequency is more than 10 – /br/, whose frequency is 12 and /hr/, whose frequency is 14.

Next in order is the type *fricative+nasal* (Table 5.37), with 10 clusters and 37 cluster tokens. There is no high-frequency cluster of this type. However, this type has more clusters and a clearly higher number of tokens than the type of the opposite order (see Table 5.26), which is against the expectation that those clusters are not favoured where sonority increases towards the syllable edge.

All the remaining clusters have low frequencies. These include the sequence *stop+nasal* (Table 5.32), the sequences with an affricate as a member (Tables 5.35 and 5.36), as well as the sequences with a glide as a member (Tables 5.34 and 5.39).

The number of clusters and cluster frequencies of different manner subclass sequences given above in Tables 5.12-5.39 are summarised in Table 5.40.

Table 5.40: *A summary of the cluster and token totals of different manner subclass sequences presented in Tables 5.12-5.39*

A. Clusters with members from the same manner subclass					
1. obstruent+obstruent			**2. sonorant+sonorant**		
	cluster	**token**		**cluster**	**token**
stop+stop	18	33	**nasal+nasal**	1	3
affricate+affricate	0	0	**liquid+liquid**	0	0
fricative+fricative	19	33	**glide+glide**	0	0
Total	**37**	**66**	**Total**	**1**	**3**

B. Clusters with members from different manner subclasses					
1. obstruent+obstruent			**2. sonorant+sonorant**		
	cluster	**token**		**cluster**	**token**
affricate+stop	3	5	**liquid+nasal**	3	18
stop+affricate	0	0	**nasal+liquid**	2	9
fricative+stop	21	134	**glide+nasal**	2	7
stop+fricative	21	43	**nasal+glide**	0	0
fricative+affricate	2	2	**glide+liquid**	2	14
affricate+fricative	3	3	**liquid+glide**	0	0
Total	**50**	**187**	**Total**	**9**	**48**

3. sonorant+obstruent in both orders								
(a) clusters with stops			**(b) clusters with affricates**			**(c) clusters with fricatives**		
	cluster	**token**		**cluster**	**token**		**cluster**	**token**
nasal+stop	12	71	**nasal+affricate**	1	9	**nasal+fricative**	8	20
stop+nasal	9	16	**affricate+nasal**	1	3	**fricative+nasal**	10	37
liquid+stop	13	98	**liquid+affricate**	2	10	**liquid+fricative**	13	63
stop+liquid	13	58	**affricate+liquid**	1	4	**fricative+liquid**	12	55
glide+stop	5	8	**glide+affricate**	0	0	**glide+fricative**	6	9
stop+glide	3	3	**affricate+glide**	0	0	**fricative+glide**	3	4
Total	**55**	**254**	**Total**	**5**	**26**	**Total**	**52**	**188**

Section A in Table 5.40 shows the statistics of clusters whose members come from the same manner subclass. Since members of the same manner subclass were assumed to have the same degree of sonority, clusters in Section A are interpreted to be sequences where the sonority level remains the same. Section B deals with clusters whose members come from different manner subclasses. Since different manner subclasses were assumed to represent different sonority levels, clusters in Section B are interpreted to be sequences where the sonority level decrease or increases, depending on the consonant order. The clusters in section B are presented in pairs with the same subclasses as members in each pair, but in reversed order. The first line in each pair shows a sequence of decreasing sonority, which is expected to be the favoured case.

When looking at the statistics of the different sequence types, it is good to recall that there are eight stops and eight fricatives, so that all sequence types which consist of these subclasses are mutually comparable. Furthermore, affricates, nasals and liquids each have two consonants, so that clusters consisting of those subgroups are mutually comparable, and so on.

We have seen before that with the exception of stops and fricatives, consonants from the same manner subclass do not usually form clusters and, when they do, their cluster frequencies are low (Section A of Table 5.40). When consonants come from different subgroups within the same major manner class (Table 5.40, Section B, l and 2), there is some increase in the number of clusters, and a clear increase in cluster tokens, even though there are still some sequences that do not occur. In Bl, the cluster type *fricative+stop* has a particularly high token frequency, and in B2, clusters with a liquid have higher token frequencies than other combinations. The pairwise comparison of the cluster types shows that the clearest difference lies in cluster token frequencies, which are practically always higher – sometimes dramatically so – in the pair of the first line, where the sonority decreases.

The last three groups of Table 5.40 (Section B 3 (a)-(c)) present sonorant/obstruent combinations in both orders. We see, first, that clusters with a stop as a member (Section 3 (a)) have higher token frequencies than other cluster types in 3, whereas the number of clusters with a stop as a member is almost the same as when a fricative is a cluster member (Section 3 (c)). Since stops and sonorants are at the opposite ends of the sonority scale of consonants, we can see that in the sonorant/obstruent sequences, those clusters where members differ more in their sonority are preferred. The pairwise comparison shows that clusters with a decreasing sonority have a higher token frequency, with one exception that has already been noted before. In section 3 (c), *nasal+fricative,* a cluster type with a decreasing sonority, has a lower token frequency than the opposite order, where the sonority increases. This goes against the expectation that clusters with a decreasing sonority are preferred, i.e. they are more frequent. We can see another difference in fricatives compared to stops: in sonorant/stop clusters, there is a great difference in token frequencies between clusters with decreasing sonority and clusters with an increasing sonority, whereas the token frequency of the sequence *liquid+fricative* is not much higher than in the opposite order. But since fricatives are next to sonorants on the sonority scale (and further from liquids than from nasals), the figures support the idea that those clusters that differ more in sonority are favoured more.

The table shows a clear preference to clusters with decreasing sonority, compared to clusters where the sonority remains the same, or increases. The following figures based on Table 5.40 further confirm this generalisation:

Clusters of decreasing sonority:	93 clusters	468 tokens
Clusters of increasing sonority:	78 clusters	235 tokens
Clusters with unchanging sonority:	38 clusters	69 tokens
Total	209 clusters	772 tokens

Finally, we can have a look at the ten most preferred clusters in Appendix 9. They are as follows:

/st/	33	/rs/	17
/ng/	26	/χt/	17
/rd/	26	/ft/	15
/št/	22	/hr/	14
/nd/	22	/rz/	14

The list shows that the two tendencies, dissimilarity and decreasing sonority, are clearly at work. All of the clusters consist of consonants from different manner subclasses: five clusters are of type *sonorant+non-continuant* obstruent (i.e. /ng/, /rd/, /nd/, /rs/, /rz/), four are of the type *continuant obstruent+non-continuant obstruent* (*fricative+stop*, namely, /st/, /št/, /χt/, and /ft/), and one cluster, /hr/ is of the type *continuant obstruent+sonorant*. As to the sonority, all other clusters except /hr/ have a less sonorous consonant as the second member. Thus, the ten most frequent clusters obey the tendencies almost without an exception. Moreover, we see that all the clusters have at least one dental, and half of them consist of dentals only.

5.6. A survey of gaps

Let us come back to Table 5.6 – given here as Table 5.41 – and look at it from the point of view of gaps. The table shows that certain consonants have restrictions in the C_1 position. Some of them, as listed in (1), do not occur as the first member in any cluster, and some others, as listed in (2), do not cluster if C_2 is occupied by an obstruent. Likewise, there are consonants that have restrictions in the C_2 position where some consonants, as mentioned in (3), do not occur in C_2 if the first cluster member is an obstruent, and neither of the affricates occurs in C_2 when the C_1 consonant is a non-continuant, i.e. a stop or a fricative. This is mentioned in (4) below. The table shows that:

(1) /p/, /č/ and /ž/ do not occur in C_1 in any cluster.
(2) /g/ and /v/ do not occur in C_1 with an obstruent in C_2.
(3) /p/, /č/ and /ž/ do not occur in C_2 with an obstruent in C_1.
(4) The affricates do not occur in C_2 with a non-continuant consonant in C_1.

We saw in the previous chapter that existing clusters show a tendency towards dissimilarity of cluster members. This suggests that gaps may be found in slots where cluster members would be identical or similar. This is indeed the case. First, identical consonants do not cluster (5), i.e. the diagonal is empty. Second, consonants that differ only in voicing (i.e. that have the same place and manner of articulation) tend not to cluster (6). There is only one exception to this, namely, /fv/ (in the syllable /ʔäfv/[50]), which occurs only in one syllable in the DBS:

[50] In Modern Persian there is a strong tendency to replace the final /v/ in the example /ʔäfv/ with a prolonged nuclear vowel, /ä/ in pronunciation.

(5) Clusters of identical consonants do not occur.

(6) Consonants with the same place and manner of articulation, differing only in voicing, do not cluster.

We also see that certain manner subclasses do not cluster together. Thus, sibilants do not form clusters together (7), and neither do liquids (8):

(7) Sibilants do not cluster together.

(8) Liquids do not cluster together.

Consonants from the same manner subclass whose places of articulation are in the same overall area may also avoid clustering. Thus, stops in the back area (i.e. in the velar, uvular and glottal areas[51]) do not cluster together (9), and neither do the back area fricatives /χ/ and /h/ (10). There is also a strong tendency for the back area fricatives not to form clusters with the back area stops (11). There is only one exception, the cluster /ɢh/, but it only occurs in one syllable of the DBS:

(9) Back area stops (i.e. velar, uvular and glottal) do not cluster with each other.

(10) Back area fricatives (i.e. uvular and glottal) do not cluster with each other.

(11) Back area stops and fricatives tend not to cluster with each other (in either order; note the one exception above).

[51] Glottals are not really [+back] consonants. I use the term *back area* to cover both the [+back] consonants proper and the glottal ([+low]).

Table 5.41 (= Table 5.6): *CC clusters in the CVCC syllables of the DBS*

$C_1\downarrow$ $C_2\rightarrow$	p	b	t	d	k	g	G	ʔ	č	ǰ	f	v	s	z	š	ž	χ	h	r	l	m	n	y	
p																								0
b			+	+	+		+	+					+	+	+		+	+	+	+		+		13
t		+			+		+	+			+							+	+	+	+	+	+	11
d							+					+	+					+	+	+	+			7
k			+										+						+	+	+	+		6
g																				+				1
G		+	+	+				+			+	+	+	+	+			+	+	+	+		+	14
ʔ		+	+	+							+		+	+	+				+	+	+	+	+	12
č																								0
ǰ		+		+				+				+		+				+	+		+			8
f			+				+	+				+	+	+	+		+		+	+		+	+	12
v																			+					1
s		+	+	+	+		+	+		+	+						+	+	+	+	+	+		14
z		+		+			+	+		+	+	+							+	+	+	+		11
š			+	+	+	+	+				+	+							+		+	+	+	11
ž																								0
χ			+										+	+	+				+	+	+			7
h			+	+								+	+	+	+				+	+	+	+	+	11
r		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+	+		19
l		+	+	+	+		+	+			+	+	+				+	+			+			12
m	+	+	+	+			+	+					+	+	+				+	+		+		12
n		+	+	+	+	+		+		+	+		+	+			+	+						12
y		+	+	+	+			+			+		+	+	+		+	+	+	+	+	+		15
	1	11	14	13	8	3	10	12	1	4	10	9	13	11	9	1	7	10	16	14	14	12	6	209