

The Vowel Harmony Patterns of Kazak Vocabulary

Ruiqing Xin, Yonghong Li, Hongzhi Yu

*Key Lab of China's Ethnic Languages Information and Technology of Ministry of Education,
Northwest Minzu University, Lanzhou, China
ruiqingxin@163.com*

Keywords: vowel harmony patterns; Kazak vocabulary; corpus of Kazak vocabulary

Abstract: With the program written on the platform of Mat lab, data in the corpus of Kazak vocabulary are computed and the vowel harmony patterns of Kazak words haven been drawn and compared with the vowel harmony patterns put forward by Geng Shimin. Results of the experiment indicate that all the four back vowels and the five front vowels can be grouped in harmony patterns respectively, which is different from what Geng Shimin has drawn.

1. Introduction

Vowel harmony is a common linguistic phenomenon which was firstly discovered in Fenland from the Ugric language family and Turkish from Altaic language family. Kazak is a member of Altaic language family with the typical property of vowel harmony.

The up to date research literature has indicated 8 classes of vowel harmony in line with the basic features of harmony phenomena, which are palatal harmony, tongue root position harmony, tongue height vowel harmony, rounding vowel harmony, nasal vowel harmony, pharyngeal vowel harmony, tenseness vowel harmony, and retroflex vowel harmony [1]. The vowel harmony of Kazak is palatal harmony. In the system of palatal harmony, all vowels in the language are divided into two groups—front vowels and back vowels. Thus, according to the principle of vowel harmony, in a certain word, all vowels should be front vowels or back vowels.

In any language, the number of vowels must be more than 2, and in some words there may be several different vowels. However, due to the limited number of vowels in any language, the vowel harmony patterns, which are the collocation patterns of vowels in words, must be limited and following a certain rule.

The findings of vowel harmony patterns can offer a higher and wider sight at the phonetic system and great help for the study of individual phonetic properties.

2. Vowel System of Kazak

Kazak is a common language for all Kazakh people in the world. At present there are mainly two writing systems of Kazak. One is in Cyrillic for Kazakhstan, the other one is in Arabic for China.

Although the writing systems are different, the pronunciations of the phonemes are almost the same, and all Kazak people can communicate well. In China, there are altogether 33 phonemes in Kazak, among which 24 are consonants and 9 are vowels. According to Geng Shimin & Li Zengxiang [2], the 9 vowels include 5 front vowels and 4 back vowels, which are listed in Fig. 1.

	Unrounded		Rounded	
	Wide vowels	Narrow vowels	Wide vowels	Narrow vowels
Back vowels	ɑ	ə	o	u
Front vowels	æ、e	i	ø	y

Figure 1. Kazak vowels classification by Geng Shimin

There are three main items for the analysis of the vowels. They can be divided into rounded vowels and unrounded vowels in term of lip-rounding, wide vowels and narrow vowels in term of mouth opening (i.e. tongue height), and front vowels and back vowel in term of tongue region. The contrast between front vowels and back vowels is the most striking feature of Kazak vowels[3]. This is also the basis of palatal harmony.

3. The Vowel Harmony Patterns of Kazak Vocabulary

Present researches have found out some vowel collocation patterns in Kazak. Geng Shimin [4] put forward the following 9 collocation patterns in 1989: [a] can only be followed by [a] or [ə]; [ə] can only be followed by [a] or [ə]; [e] can only be followed by [e] or [i]; [i] can only be followed by [e] or [i]; [æ] can only be followed by [e] or [i]; [o] can only be followed by [a] or [ə]; [ø] can only be followed by [i] or [ə]; [u] can only be followed by [a] or [ə]; [y] can only be followed by [i] or [e]. These patterns can be listed in table I.

Table I. Vowel collocation patterns of geng shimin

The First Vowel	Possible Following Vowels
[a]	[a] or [ə]
[ə]	[a] or [ə]
[e]	[e] or [i]
[i]	[e] or [i]
[æ]	[e] or [i]
[o]	[a] or [ə]
[ø]	[i] or [ə]
[u]	[a] or [ə]
[y]	[e] or [i]

Cheng Yanyan [5] summarized the lip-rounding vowel harmony patterns into 6 pairs: ø-ø, u-u, y-y, o-u, ø-y, y-ø. Lip-rounding vowel harmony can only be found in spoken Kazak not in written Kazak.

This experiment was based on a large corpus of Kazak vocabulary, and was aimed to find out all the possible vowel collocation patterns in word formation.

3.1 Preparation for the Corpus of Kazak Vocabulary

The very beginning of the corpus of Kazak vocabulary is the Internet Kazak Corpus—Kazak-Chinese Dictionary uploaded by Li Jing [6]. All words and their affiliated information from *Kazak-Chinese Dictionary* compiled by Nurbek Abken [7] are concluded in the corpus. To guarantee the validity of the research, each entry of word in the corpus was examined according to *Kazak-Chinese Dictionary*.

In Kazak there are many words borrowed from other languages, such as Arabic, Persian, Russian, Chinese, and Indo-European languages. Because some words borrowed from Indo-European languages are pronounced according to the original words, and not following the rule of vowel harmony, it is very necessary to filter out the Indo-European borrowed words firstly to guarantee the validity of the experiment and provide convenience for further data processing.

However, since Kazak in China is written in Arabic symbols, there is a close relationship between Kazak and Arabic and even Persian, and words borrowed from Arabic and Persian have been assimilated into Kazak completely, and all the borrowed words follow the rule of vowel harmony. Even some words borrowed from Arabic and Persian are closely relevant to Kazakhs' daily life and more commonly used than some Kazak words. Hence, it is wise to include words borrowed from Arabic and Persian as traditional vocabulary. Also, words borrowed from Russian and Chinese have been assimilated into Kazak since the beginning, but considering the large gap between the languages, these words are classified as borrowed words, and should be shifted out in this experiment.

The identification of borrowed words is a hard job. First, with the information from *Kazak-Chinese Dictionary*—[阿] is for Arabic, [波] is for Persian, [俄] is for Russian and [汉] is for Chinese—words borrowed from Arabic, Persian, Russian and Chinese were marked. Then taking reference of *Kazak Etymologicon* the origins of the words were found out, and then with the help of comparison between Kazak and English, most words borrowed (i.e. words sharing similar pronunciations and definitions in Kazak and English) from Indo-European have been extracted and marked.

With the filtering capability of Excel, demanded words were extracted and others were excluded.

3.2 Data Processing

Programs were written on the platform of Mat lab and carried out step by step for the processing of the data in the corpus.

Firstly, each word was attributed in line with the affiliated information and features, such as words with indicators for compound words, words affiliated with numbers for types and so on. The attributes and the words themselves were listed in different columns, and thus, without affiliated information words were more manageable for further computing.

Secondly, compound words were divided into front words and back words and listed separately in columns. Other words without an obvious indicator for compounding, such as space or hyphen, were listed in the same column as front words. At the same time, the marker for front vowels “ء” was elicited from front words and back words respectively, and attributed in separate columns.

Thirdly, Kazak words were converted to IPA in line with the check list between Kazak phonemes and IPA (see table II) and the pronunciation principles of Kazak: When the front vowel marker is put in the front of a word, that is to say all of the vowels in the word should be pronounced as front vowels. When there is [e], [g] or [k] in a word, the letters ا [a], ع [ə], و [o], ؤ [u] should be pronounced as their front vowel counterpart respectively even without a front vowel marker. When the letters “ۆ” and “ي” are after vowel letters, they are pronounced as the semi-vowels [w] and [j],

while when they are after consonant letters, and there is no vowels after them, they are pronounced as [uw] and [əj], or [yɥ] and [ɨj] with a front vowel marker.

Table II. Check list between kazak phonemes and ipa

	Kazak	IPA		Kazak	IPA
Front vowels	е	e	Voiced consonants	б	b
	ә	æ		д	d
	ә	ø		г	g
	ү	y		қ	q
	і	i		с	ɕ
Back vowels	а	ɑ		з	z
	о	o		ғ	ɣ
	у	u		м	m
	ә	ə		н	n
Voiceless consonant	п	p		Approximant	л
	т	t	ң		ŋ
	к	k	р		r
	қ	ʧ	й		j
	ғ	f	ұ		w
	с	s			
	ш	ʃ			
	х	x			
	һ	h			

Finally, words were divided into syllables according to the rules for syllable division [8]: If there is only one consonant between two vowels, the consonant should be grouped into the later syllable; if there are two consonants between two vowels, the one consonant should be grouped into the former syllable and the other into the later syllable respectively; if there are three consonants between two vowels, the first two consonants should be grouped into the former syllable and the other consonant into the later syllable. After the division of the syllables, vowels in each syllable of the word were extracted. Then, the vowels were summed together, and the vowel collocation patterns have been extracted. The sample of the final data processing results is shown in Fig. 2.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
WN	KAZAK	WC	Chinese	FW	S1	S2	S3	S4	S5	S6	S7	S8	V1	V2	V3	V4	V5	V6	V7	V8	VS
N00001	а		哈萨克字母表中的第一个字母	a	a								a								a
N00003	аба	noun.	<名> (圈牲畜的) 围栏	aba	a	ba							a	a							aa
N00004	аба	noun.	<名> 披风, 斗篷。	aba	a	ba							a	a							aa
N00005	аба	echo.	(狗) 狂吠。	aba	a	ba							a	a							aa
N00006	абадан	noun.	<名> (阿) (古) 结实, 健壮; 力大如牛, 大力士。	abadan	a	ba	dan						a	a	a						aaa
N00007	абаға	noun.	<名> 大木箱、大箱子。	abaға	a	ba	ға						a	a	a						aaa
N00008	абағадая	adj.	<形> (1) 笨重的, 大而粗笨的, 不灵活的	abaғадая	a	ba	ға	daj					a	a	a	a					aaaa
N00010	абағуырлы	adj.	<形> 带灯罩的, 有灯伞的, 有灯罩子的	abaғуырлы	a	ba	ғуыр	лы					a	a	u	ə					aaue

Figure 2. Sample of the final data processing results for extraction of vowel collocation patterns
 WN=Word Number, WC=Word Class, FW=Front Word, S=Syllable, V=Vowel, VS=Vowel Sum

3.3 Exactraction of Vowel Harmony Patterns

The vowel collocation pattern of every word is actually the vowel harmony pattern. Data in the

column of VS (Vowel Sum) relevant to front words and back words were elicited for categorization to draw the types of harmony patterns and the quantitative distribution.

First, data relevant to front words were extracted, repeated words were cut out, and harmony patterns were drawn for categorization. All together, 883 kinds of patterns were available. Part of the results can be found in Fig.3.

1	2	3	A	B
+	211	a 计数		207
+	1976	aa 计数		1764
+	3982	aaa 计数		2005
+	4347	aaaa 计数		364
+	4358	aaaaa 计数		10
+	4361	aaaaaa 计数		2
+	4363	aaaaaa 计数		1
+	4366	aaaaau 计数		2
+	4395	aaaaa 计数		28
+	4398	aaaaaa 计数		2
+	4400	aaaaaa 计数		1
+	4403	aaaaaa 计数		2
+	4430	aaaau 计数		26
+	4787	aaa 计数		356
+	4822	aaaa 计数		34
+	4824	aaaaa 计数		1
+	4826	aaaaau 计数		1
+	4831	aaaaa 计数		4
+	4835	aaaau 计数		3
+	4877	aaaa 计数		41
+	4884	aaaaa 计数		6
+	4886	aaaaa 计数		1

Figure 3. Part of the results for front words vowel harmony patterns extraction

Then, likewise, harmony patterns of back words were drawn and all together 135 kinds of patterns were available.

Finally, all harmony patterns of front words and back words were put together for categorization, and all together 886 kinds of patterns were available.

4. Discussion of the Experiment Results

The vowel collocation patterns achieved in this experiment have been listed in table III. Comparison between the experiment results and the vowel harmony patterns put forward by Geng Shimin was made.

Table III. Vowel collocations of this experiment

The First Vowel	Possible Following Vowels
[a]	[a] [ə] [u] [o]
[ə]	[a] [ə] [u] [o]
[e]	[æ] [e] [i] [ø] [y]
[i]	[æ] [e] [i] [ø] [y]
[æ]	[æ] [e] [i] [ø] [y]
[o]	[a] [ə] [u] [o]
[ø]	[æ] [e] [i] [ø] [y]
[u]	[a] [ə] [u] [o]
[y]	[æ] [e] [i] [ø] [y]

Results indicate that all the four back vowels and the five front vowels can be grouped in harmony pattern respectively, which is different from what Geng Shimin has drawn.

For example, after the vowel [a], possible vowels are [a] [ə] [u] [o] instead of just [a] and [ə], like the words “ابۇير [abujər]” and “اتوي [atoj]”. Enough evidences are available to testify the rest of the experiment result.

The differences between the results maybe explained by the limited language data collected in the 1980s or the research technologies. Anyway, more experiments and researches are really needed for the collection of Kazak vowel harmony patterns.

Acknowledgment

This paper is available thanks to the support of the projects: Northwest Minzu University Project Supported by the Fundamental Research Funds for the Central Universities: A Study of Vocabulary from Kazak and English Sharing Similar Pronunciations and Definitions (Project No.: 31920180064) and National research project of the National Committee of People’s Republic of China (post financing project): The Study of Languages Reflecting the Cultural Fusion along the Belt and Road (Project No.: 2017-GMH-015).

References

- [1] Li Bing, *Vowel harmony of Altaic language family*, Beijing: The Commercial Press, 2013, pp.25-26.
- [2] Geng Shimin & Li Zengxiang, *An introduction to Kazak*, Beijing: The Ethnic Publishing House, 1985, pp.3.
- [3] Xin Ruiqingxin, “The pattern of tongue positions and properties of Kazak vowels”, *IOP Conference Series: Earth and Environmental Science*, Vol. 69, Jun. 2017, 02173.
- [4] Geng Shimin, *Modern Kazak grammar*, Beijing: Publishing House of the Central University for Nationalities, 1989, pp. 19-20.
- [5] Cheng Yanyan, *Lexicology of modern Kazak*, Beijing: The Ethnic Publishing House, 2000, pp. 17.
- [6] Li Jing, *Internet Kazak corpus — Kazak-Chinese dictionary*, <http://vdisk.weibo.com/s/taoCn8PUHOqk4>, 2015.
- [7] Nurbek Abken, *Kazak-Chinese dictionary*, Beijing: The Ethnic Publishing House, 2014.
- [8] Geng Shimin & Li Zengxiang, *An introduction to Kazak*, Beijing: The Ethnic Publishing House, 1985, pp.10.