



THE PRINCIPLES OF COMPOSITION OF THE CORPUS OF LACUNAR UNITS IN UZBEK DIALECTS

Abdurahmanova Sayyora Bozorboy qizi

Basic doctoral student, Tashkent State University of
Uzbek Language and Literature named after Alisher Navoi
sayyoraabdurahman@gmail.com

Abstract:

The article talks about the principles of creating a corpus of lacunar units in Uzbek dialects. At the same time, it will reveal the subtle meanings underlying these lacunar units, and the alternative that is used instead of the concept expressed by these lacunar units.

Keywords: lacuna, platform, corpus, dialect, principle, linguistic culture.

Due to independence, the attitude towards spiritual values and the art of speech has changed radically in every aspect. The opportunities given by independence, that is, the new era, required a different and new approach in all aspects. The great change and renewal in our spiritual life has opened a wide way for comprehensive study and research of our tangible and intangible heritage. This is natural, of course.

Renewal, development or fundamental change of any system is connected with the change of national spiritual and educational thinking. As a result of interdisciplinary integration, fundamental changes are taking place in all areas and fields. As a result of such changes in linguistics, new research objects are emerging. It would not be wrong to say that the science of computer linguistics, which is one of the modern directions of linguistics, is one of the most demanding fields today.

We can include a number of words that occur only in the Zomin dialect in the corpus of lacunae.

Livestock-related *chabara* (a closed, fenced, off-yard enclosure for keeping cattle or sheep unencumbered), or *nova/nava* (cattle or sheep). The words *zilar* (water tank) are combined under the common words *kora* or *akhur*. In addition, *black kilang/black kilan* or *kilan*, representing natural phenomena (the cloudiness of the air after snowfall and the melting of the snow at the same time) and *shivat/shibat* (the falling of very small drops mixed with fog to the ground process) words also exist in the Zomin dialect, and the words expressing these concepts are not observed in the literary language. At the same time, words such as *alakan/olakan* (palm part of the hand), *borboy* (calf), which do not have an alternative in the original Turkish or





literary language, representing the parts of the human body, are a word that refers to a person's appearance. quropli (said about people who are similar in appearance, age, gender) etc.

There is a case of using the words shibağa (the part where carrots are put in soup, zirvak) in relation to a certain part of the food preparation process. In creating the corpus of these units, we must first establish some principles. First, we will collect the necessary materials to create the body of lacunae. Lacunas in Uzbek language to form the corpus.

To create the body of these units, we must first label the uis. This process is not difficult because the phenomenon of homonymy is almost never found in lacunar units. In creating the corpus of lacunae, we use the dictionary of lacunae in the Uzbek language, dictionaries of Uzbek dialects, and sources collected in dialectological practices.

Corpus data is distributed on the basis of a license in accordance with the laws of the Republic of Uzbekistan. They also indicate the source of the materials.

In the process of creating a corpus, we create a corpus of lacunae based on several principles.

1. Planning. In this, first of all, the collected materials are gathered together, and based on this, we determine a number of goals and tasks.

What logical idea is this corpus based on;

What is the size of the data we work with;

Why is this corpus needed;

What is the level of importance;

How do I use lacunae (contextually or individually);

What topics do we work with lexical units (LB);

What tasks do we perform when creating a corpus;

The key process in creating a corpus is expressed as follows.

2. Data collection and sorting. There is another situation in this process that cannot be ignored. The main part of the information in the corpus of lacunae that we are creating is almost not found on the Internet and in non-fiction literature. Sorting out the data is somewhat difficult due to the fact that most of the lacunae are recorded from the living vernacular and these LBs are used colloquially. Attention is also required in actions related to the combination of additives. If text and audio data with gaps can be entered into this corpus into computer programs (if the corpus is oral), this corpus of the language will be very interesting and rich in content.

3. Encoding.





Another sensitive issue is file encoding. Actions, including actions performed on a computer, are convenient for working in English (more precisely, with the Latin alphabet). For this reason, many problems are related to the fact that the computer program does not recognize characters from other alphabets (for example, the Krill alphabet for the Russian language). Many examples are seen when bookmarking a web page, instead of reading any text, krill letters appear on the screen in a nonsensical sequence. This happens not only in reality known as “encodings” (English encodings), but some of them are expressed in the Russian alphabet - koi8-r or cp1251. It will not be correct to choose one of them as a template. It's also not long ago that the Unicode encoding system was introduced, which can support all the world's languages and the scripts they use, even Egyptian hieroglyphs. Not all created programs are ready to work with it yet. Any text file is saved with a unique code. Correspondingly, if a corpus analysis program considers a single code, it may not read the character set we want to use. Such a problem does not exist only in English texts.

A	B	turkum					
		C	D	E	F	G	H
so'z	izohi	ot	sifat	son	fe'l	olmosh	ravish
1. ĀBDŌL TĪLI (абдол тили)	xalq san'atkorlarining yashirin tili. Bu "til" yuzlab so'z va so'z birliklarini qamrab oladi. Masalan, paypoq maxsum – tomosha paytida tushgan puldan yashirib qoluvchi shaxs, galstuk – o'qituvchi (Xorazm, Gurlan shevasi; S.A).						
2. ĀBIRĪŪI (абиржи)	bahor paytida daryodagi suvning muz aralash oqimi (Janubiy Qoraqalpog'iston o'zbek shevasi; O'.E).						
3. ĀBŪŌŌI (абжуши)	– maska uzumdan tayyorlangan ichi danakli mayiz (Janubiy Qoraqalpog'iston o'zbek shevasi; O'.E). ishqor solingan qaynoq suvga botirib, so'ng quritilgan mayiz (Qashqadaryo, Maymanaq; N.T).						
4. ĀDRĀMKĀŖ (адрамкаш)	uchi ilmoqli katta juvoldiz (Xorazm; A.A).						
5. ĀDISĀK (адизак)	yovuz maxluq ko'rinishidagi afsonaviy kuch (Qashqadaryo, Chiroqchi; N.T).						
6. ĀFŌN - SAFŌN (афон-сафон)	– biron katta ishning yakunlanganligi, tugaganligi (Turk - kaltatoy shevasidagi ayrim so'zlar. Baxmal aravaning yuk ortiladigan qismi (O'zbek xalq shevalari lug'ati. Andijon, 1971; To'plovchi; H.A.).						
7. ĀJŪYŪZ (айжуз)	– ko'milmaydigan tok (O'zbek xalq shevalari lug'ati. Toshkent, 1971; To'plovchi; H.A.).						
8. ĀJMŌQŪI (аймоқи)	oshqozon kasalligini davolashda ishlatiladigan, karamga o'xshash bargli dorivor tog' o'simligi (O'smay, rivojlanmay qolgan; majmag' il. G'anakni almashtirish vaqtida mana bundaqa ajvaqalarni – o'smay qolgan qurtlarni olib ko'rsatdi. Oliyaxon. N. Safarov.						
9. ĀJIVĀQA (аживака)	fe'l - atvorida turg'unlik xislatlari bo'lmagan (odam) (Qashqadaryo, Uyshun; N.T). oqar suvning chiqur yeri (Buxoro, Qorako'l; A.A).						
10. ĀJNIMŌL (айнимол)	tikanli xashak (yantoq, olako'z'tikan, junak)larni o'rishda ishlatiladigan uchi ayri cho'p (Qashqadaryo, katta ariqdan chiqarilgan, ayirilgan (Qashqadaryo, Qorashina; N.T).						
11. ĀJRAM (айрам)	– Bki tuyuqning bir ul uchini to'ldirishni ma'nada, sin bilan o'rao qotirilgan, quyi qismi orasi ikki metr dan iborat yer o'zbek moslamasi (Qashqadaryo, Lager; N.T).						
12. ĀJRICŌKI (айричўқи)	1) tuyuq orasining kasallanishi; 2) ikki tuyuq (mol, qo'y, echki) (Qashqadaryo, Ko'kabuloq; N.T).						
13. ĀJRLMĀ (айрлма)	– ayitishga arziydigan, aytis bo'ladigan (Qashqadaryo, Zashkana; N.T).						
14. ĀJRTĀŌQ (айртаёқ)	aravaning (novdadan to'qilgan) yuk ortadigan qismi (Andijon; S.N).						
15. ĀJRTŪŌQ (айртуёқ)	o'ynaladigan oshiqning yutuq deb hisoblanadigan tomoni (Qashqadaryo, Dehqonobod; N.T).						
16. ĀJTŪYŪI (айтујуи)	belgilangan chiziq ustiga ost-ust qilib terilgan oshiqni sozqa bilan ma'lum masofadan urib yiqitish va marradan chiqarish, o'yin (Qashqadaryo, Toshqo'rg'on; N.T).						
17. ĀJ (аж)							
18. ĀJDI (ажди)							

a. classification of words to be placed in the corpus of lacunae

1	2	so'z	izohi	qo'shimchalar									
				-ning	-ni	-ga	-da	-dan	-m	-im	-ng	-i	
1	2	ĀBDŌL TĪLI (абдол тили)	xalq san'atkorlarining yashirin tili. Bu "til" yuzlab so'z va so'z birliklarini qamrab oladi. Masalan, paypoq maxsum – tomosha paytida tushgan puldan yashirib qoluvchi shaxs, galstuk – o'qituvchi (Xorazm, Gurlan shevasi; S.A).										
3	4	ĀBIRĪŪI (абиржи)	bahor paytida daryodagi suvning muz aralash oqimi (Janubiy Qoraqalpog'iston o'zbek shevasi; O'.E).										
5	6	ĀBŪŌŌI (абжуши)	– maska uzumdan tayyorlangan ichi danakli mayiz (Janubiy Qoraqalpog'iston o'zbek shevasi; O'.E).										
7	8	ĀDRĀMKĀŖ (адрамкаш)	ishqor solingan qaynoq suvga botirib, so'ng quritilgan mayiz (Qashqadaryo, Maymanaq; N.T).										
9	10	ĀDISĀK (адизак)	uchi ilmoqli katta juvoldiz (Xorazm; A.A).										
11	12	ĀFŌN - SAFŌN (афон-сафон)	yovuz maxluq ko'rinishidagi afsonaviy kuch (Qashqadaryo, Chiroqchi; N.T).										
13	14	ĀJŪYŪZ (айжуз)	– biron katta ishning yakunlanganligi, tugaganligi (Turk - kaltatoy shevasidagi ayrim so'zlar. Baxmal tumani; H.A.).										
15	16	ĀJMŌQŪI (аймоқи)	aravaning yuk ortiladigan qismi (O'zbek xalq shevalari lug'ati. Andijon, 1971; To'plovchi; H.A.).										
17	18	ĀJRTĀŌQ (айртаёқ)	– ko'milmaydigan tok (O'zbek xalq shevalari lug'ati. Toshkent, 1971; To'plovchi; H.A.).										
19	20	ĀJRTŪŌQ (айртуёқ)	oshqozon kasalligini davolashda ishlatiladigan, karamga o'xshash bargli dorivor tog' o'simligi (Qashqadaryo, Dehqonobod; N.T).										
21	22	ĀJIVĀQA (аживака)	O'smay, rivojlanmay qolgan; majmag' il. G'anakni almashtirish vaqtida mana bundaqa ajvaqalarni – o'smay qolgan qurtlarni olib ko'rsatdi. Oliyaxon. N. Safarov.										

b. separating the words to be placed in the corpus of lacunas according to the acceptance of suffixes.





They are read, coded and analyzed normally. Body label (label, symbol, annotation). You can define your own text and add special characters to it. Markup helps you to identify a specific place in the text. Punctuation is the compactification of texts using punctuation marks.

References:

1. Xolmanova Z. O'zbek shevalaridagi lakunalar lug'ati
2. Abdullayeva O. O'zbek tilidagi internet matnlari korpusini yaratish. f.f.f.d., PhD dissertatsiyasi
3. Qozoqboy Yo'ldoshev, Bosim To'ychiboyev Zominning til qomusi
4. Z. Xolmanova, O. Saidahmedova, O. Nurullayeva Lingvokulturologiyaga oid tushunchalar tadqiqi – Toshkent 2018.

