JAKUB ŁUKASIK
Jagiellonian University in Kraków
jak.mic@wp.pl
ORCID: 0000-0002-5820-2356

THE ETYMOLOGY OF THE TURKIC *EV*, *EB*, *ÜJ* ETC. 'HOUSE'[1]

Abstract

This paper collates and reviews previously proposed etymologies of the Turkic word *ev*, *eb*, *üj* etc. 'house'. Moreover, as the etymology of this lexeme is still uncertain, it aims to once again analyze the available material and attempts to establish a convincing etymology. Another important aim of the paper is to clearly demonstrate which issues connected with the topic can be considered established and which require further study.

Introduction[2]

The Turkish *ev*,[3] together with its cognates, constitute the main Turkic word for 'house'. It is present in all the Turkic languages, with the exception of Chuvash, Yakut and Dolgan.[4] Although in some languages it has ceased to be the primary designation of 'house', it is still present, albeit in different, more specified functions, such as in the Tuv. *öγ* 'yurt' (Tenišev 1968: 330).

Much has been written about its origin, but most contributions have been limited to single paragraphs in works primarily focused upon different subjects. The

---

[3] From this point onwards, it is written in a transcription consistent with the other languages, that is as *äv*.

[4] I have been unable to obtain data on Salar.

single exception, an article by Polivanov (1934), was published in Uzbekistan and today is unfortunately not readily available to western scholars. Despite the interest in this lexeme, modern authors still view its etymology as being uncertain (Tatarincev IV 354–355).

The purpose of this paper is to: 1. collate and review the previous etymological propositions, including lesser-known contributions (section 1); 2. perform a detailed analysis of the attested forms of the lexeme, both ancient and modern, in order to provide a basis for its reconstruction (sections 2 and 3); 3. attempt to offer a systematic etymological analysis and reconstruction (section 4.2); 4. answer further questions regarding this lexeme that have arisen during the analyses (sections 4.1, 4.3 and 4.4).

1. A comprehensive literature overview

Below, is a comprehensive review of the proposed etymologies of the lexeme in question, in chronological order. Each etymology is presented together with its author's argumentation. The discussion was reduced to the minimum, as I will present my own detailed etymological analysis in the latter part of the paper. The review is followed by Table 1, in which each reconstruction is presented in tabular form.

1.1. W. Radloff (1882)

Radloff (1882: 184, 186) reconstructs *äp, from which he derives äv, äw, äj, öj. Although, it is unclear whether he thinks that they constitute a development chain or whether they all emerged directly from *äp, he seems to suggest the latter, stating that ü evolved from äv, while üj and üg developed from öj (again leaving it unsaid whether öj > üj > üg or öj > üj, üg). Both interpretations of Radloff's proposal are presented below (for readability's sake, doubts concerning the development of the forms with a labial vowel are not included):

   *äp > äv (> ü), äw, äj, öj (> üj > üg), or:
   *äp > äv (> ü) > äw > äj > öj > üj > üg

1.2. V. Grønbech (1902)

V. Grønbech (1979: 64–65),[5] contrary to the other etymologists, views the development of äb as being regular. He does not construct long development chains, limiting himself to stating that he believes the PT form to be *äb. He also presents a number of lexemes which he suggests display the same development: Ott. säv- 'to love',

---

[5]   The difference in dating between this reference and the title of the section is due to the fact that I cite the English translation of Grønbech's work, which was published much later than the original.

dävir- 'to turn up and down', and *čävir-* 'to turn'. Unfortunately, for the latter two the only cognate he cites comes from Chuvash (in his transcription: *tawër-*, *śawër-*), which does not have a reflex of **äb* (it does, however, have a derivative, *avlan-* 'to get married').

1.3. M. Hartmann (1904)

Hartmann (1904: 176) critiques Radloff's etymology, stating that the **äp* form (written inconsistently, sometimes with a broad *ä*, and sometimes with a narrow *e*) cannot be the original because in "Kutadgu Bilig" there is the form *ebi*. Hartmann argues that *eb* must be the older form, and that it became devoiced while "standing on its own or before a hard consonant".[6] The opinion that forms with *b* are older than those with *p* is shared by most researchers. Nevertheless, Hartmann's argumentation lacks a certain credibility, as it could equally be argued that the consonant in **äp* underwent voicing in the intervocalic position. Subsequently, he puts forward his own proposition, which is as follows:

   *eβ > eb (> ep), ev, öj, üj

His argumentation is somewhat vague in that he does not explain his reasoning for reconstructing a narrow *e* (unless this is unfortunately the result of his inconsistent notation, *vide supra*). He reconstructs *β* as a "middle ground" of sorts between *b* and *v*. The forms *eb* and *ev* are supposed to have developed simultaneously, the former in "East Turkic"[7] and the latter in "South Turkic"[8] (= Oghuz), and he fails to specify whether *öj* and *üj* come directly from **eβ*, or from one of the later forms.

1.4. E.D. Polivanov (1934)

Polivanov (1934, quoted after ÈSTJa I 514) reconstructs an initial narrow *e*. Sadly, I was unable to access the full text of Polivanov (1934) during the preparation of this paper, and thus his reasoning for such a reconstruction [omitted by Sevortjan (ÈSTJa I 514)] remains unknown to me. However, in all likelihood, the reason was the existence of forms with both high and low vowels.

   He then suggests the following development:

   *eb > *ep > *ew > *eü > *oj, öj.

Polivanov finds the change **ew > *eü > oj* analogous to the Ger. *eu > oi*. A somewhat problematic part of this etymology is the scarce attestation of the forms *eü* and *oj*. Sevortjan (ÈSTJa I 514) adduces one occurrence of *oj*, in Uzbek dialects, where it is most likely secondary and comes from **öj* (cf. section 2.6). The form *eü* is found in a single, poorly attested dialect of Siberian Tatar (R I 671). Therefore, it is difficult to

---

6   "Eb musste alleinstehend oder vor harten Konsonanten ep werden" (Hartmann 1904: 176).
7   *Osttürkisch* (Hartmann 1904: 176).
8   *Südtürkisch* (Hartmann 1904: 176).

accept these forms as early or common in the Turkic languages. It is more probable that Polivanov simply deemed them hypothetical steps in his development chain in order to better ground his analogy to German, although it is still unclear why he insisted on *eü, which seems redundant. Perhaps these issues are explained more clearly in his article.

1.5. W. Bang (1936)

Bang (1936: 43) reconstructs the following string of phonetic developments:

*äb, *ib > *äv, *iv + -i (poss. suf.) > *ävi, *ivi > *äwi, *iwi > *äwü, *iwü > *öwü, *üwü > *ögü, *ügü (> ög, üg), *öjü *üjü > öj, üj.

The change of *äb, *ib into *äv, *iv is accepted by Bang, without further explanation. Further changes are considered to be the result of the addition of the possessive suffix -i. The v, after finding itself in the intervocalic position changed into w, which was consequently "weakened", leading to the emergence of a hiatus, later eliminated by the addition of g or j. Sevortjan (ÈSTJa I 514) writes, that: "[…] here -g […] is interpreted by Bang to come from -j",[9] but in fact this is not what Bang wrote, as he believes the forms with g and j to have developed simultaneously. The labialization of the vowel is not explicitly explained, but from the development chain cited above, it can be deduced that Bang proposed two changes. Firstly, the intervocalic w labialized the suffix, and later the root vowel was assimilated to the suffix. If this were the case, it would be a rather unusual example of labial harmony working backwards. He suggests *suγγ ≪ *sub 'water' as a potential parallel example.[10] The monosyllabic forms ög, üg, öj, üj are supposed to have later been abstracted from the suffixed forms, and the forms of the type ū̃ are the result of the contraction of *ügü. The forms öp, öb. öv, and öm developed directly from *äb:

*äb > *äb, *äv, *äp, *äm > öb, öv, öp, öm.

The rather peculiar feature of this etymology is the reconstruction of two alternating forms, one with a broad vowel and the other with a narrow vowel. This allows Bang to explain the differences in the height of the vowel observed in modern languages without having to decide which form was the older. However, the emergence of this alternation still requires an explanation, which Bang fails to provide.

1.6. G.J. Ramstedt (1949)

Ramstedt (1949: 56) concurs that the reconstructed form should be *eb. but fails to discuss the etymology further. His unique contribution is connected to the origin of

---

[9]   "[…] Здесь -г также рассматрается В. Бангом в качестве производного от -й."

[10] In fact it is the opposite, as the paragraph in Bang's work is mainly concerned with sub, and äb is only included as a supporting example.

the forms *öγ*. Ramstedt himself only comments on a single form, which he ascribes to Koibal: *öγ* 'Erbteil, Haus'. He believes that it does not come from the PT **eb*, but is a borrowing from the Mongolian *öb* 'inherited land or possessions, domain'. While Ramstedt only writes about Koibal, I see no reason not to extend this explanation to all the other instances of the same form.[11] Tatarincev (Tatarincev IV 354–355) does exactly this when critiquing Ramstedt (cf. section 1.11).

Unfortunately, I have been unable to locate a Koibal form of this shape, and Ramstedt also fails to provide a source for it. Furthermore, he does not present any additional argumentation. This issue will be dealt with in more detail in section 4.1.1.

1.7. M. Räsänen (1949)

Räsänen (1949: 128–129) doubts the veracity of Bang's proposal and suggests an etymology similar to that of Polivanov, also utilizing the German analogy, albeit without the redundant steps noted above:

*äb > *öv > *öj > üj.

He most likely believed the development of **äb* to be regular, as he adduces many examples of a similar evolution, but the somewhat complicated structure of his work, as well as his lack of additional commentary, makes it rather unclear. Additionally, aside from **säb-* 'to love' (> Tksh. *sev-*, Kum. *sev, söv, söj*, Uzb. *süj*, etc.), other examples do not always show the same reflexes as **äb*.

Räsänen, however, made an important remark on the origin of the forms *uγ, ū*. In his view, they do not come from the same root as all the other forms, but instead continue the PT **ūk*, and are related to the Ott. *huγ* 'hut'. This idea will be explored further in section 4.1.2.

1.8. G. Clauson (1957)

Clauson (1957: 38) proposes the following chain of developments:

*ǟv > *üv > *üw > *ū > üj

He does not support this suggestion with any concrete arguments, simply stating that the emergence of the auslaut *j* is "irregular" and "euphonic". He suggests the form *ǟv* with a long vowel comes from the 8th century, suggesting a source in the runic script. This is curious, as the runic script did not denote the length of the vowel (at least not consistently) (Róna Tas 1991: 58–59; Erdal 2004: 47–48). Additionally, other sources relating to Orkhon Turkic cite this word with a short vowel (cf. section 3.2).

---

11   Ramstedt's idea will be considered a number of times further on in the paper. To avoid repeating the same remarks, from now on I will refer to this idea as if it concerned all the examples of the form *öγ*, not only those in Koibal.

1.9. A.N. Kononov (1958)

Kononov (1958: 79–80) considers the original form to be *ob ~ *ov, from which the later forms evolved in two separate ways, one resulting in forms with a labial vowel, and the other with an illabial vowel:

oj ~ öj, uj ~ üj < *ob ~ *ov > äb ~ äp ~ äv

Unfortunately, he fails to provide any supporting arguments, simply stating that "the historical phonetics of the Turkic languages suggest another [= different than that suggested by Polivanov – JŁ] possible developmental process for this word".[12] He compares this form with sub ~ suv ~ suγ ~ su(j) ~ su 'water', and it can be deduced that this example was intended to justify the alternation of the auslaut consonant. However, it should be noted that the distribution of the reflexes is different in the lexeme sub than in äb (in terms of the auslaut), e.g. the Tksh. ev, with -v, but su with -Ø. As the vast majority of modern forms have a front vowel, the reconstruction of a back vowel raises concerns. It is possible that the reason Kononov opted for the *ob ~ *ov reconstruction was not phonetics, but rather the need that it should resemble oba 'room, family, nomad encampment', as he wished to connect the two lexemes etymologically (Kononov 1958: 79–80).

1.10. J. Hamilton (1974)

Hamilton (1974: 114) starts with *äβ > *äw, from which he directly derives forms with both a narrow and a broad vowel: *üw ~ öw. He explains the emergence of labial vowels, and the subsequent change into üj ~ öj with assimilations. The whole process can be represented as follows:

*äβ > *äw > *üw, *öw > üj, öj

Similarly to Bang, Hamilton also reconstructs two alternating forms, differentiated by the height of the vowel. Unlike Bang, however, he establishes this alternation in the middle of the development chain, not at the beginning, also leaving its emergence unexplained.

Hamilton also compares the development of this word to the OA soj 'speech, message' and the OA sojla-[13] 'to declare, to ask', which, according to his thesis, come directly from *saβ 'speech, message'. This comparison is complicated and this is not the place in which to discuss it further. It suffices to say that not everyone agrees with Hamilton (cf. e.g. KEWT 308).

1.11. B.I. Tatarincev (2008)

Tatarincev (Tatarincev IV 354–355) does not discuss the phonetic aspect of the reconstruction in depth, but instead selectively recounts a few previous proposals.

---

[12] "Историческая фонетика тюркских языков подсказывает другой возможный путь происхождения этого слова" (Kononov 1958: 79). Here, as well as in other similar instances below, the translation is my own.

[13] In Hamilton's opinion, this is an entirely different lexeme from the later Turkish söjle- 'to speak'.

He fails to share his own thoughts on the matter, generally calling the etymology "unclear". He critiques Ramstedt's idea that the South-Siberian forms with -*γ* are derived from a different root, stating that from a semantic point of view this is doubtful, and from a phonetic perspective impossible. The most substantial part of his contribution is the attempt to connect the lexeme under consideration to the verb *ävir- ~ övür-* 'to twirl, to turn'. This idea has already been expressed elsewhere in ÈSTJa I (e.g. 499–500), but Tatarincev's formulation is much more refined. For further discussion of this idea, see section 4.4.

1.12. M. Stachowski (2019)

M. Stachowski (KEWT 147) reconstructs **eb* (he maintains that Common Turkic did not distinguish the narrow /e/ from the broad /ä/, so this notation essentially means "a short vowel of the *E* type", cf. KEWT 7; M. Stachowski 1998). He does not explain the further phonetic changes, limiting himself to a comparison of the word in question with the PT **sab* 'word', and the Tksh. *söjle-* 'to speak'.

1.13. Extra-Turkic Parallels

Some researchers attempted to connect **äb* with words from different language families. Munkácsi and Hommel linked it to the Sum. *ab* (Eren 1999: 141), Polivanov (1968: 167) deemed it a borrowing from the OC **ip* 'settlement, emporium' (> Chin. *i*), and Menges (1982: 109–112) compared it to the Mkor. *ip* 'house' and the Jap. *iba ~ ipu ~ jū* id. These proposals will not be discussed here. Further bibliography on this topic can be found in Eren (1999: 141).

1.14. Summary

As the word in question consists of two phonemes, we can discuss three aspects of the etymologies presented above, namely, the vowel, the consonant, and certain common etymological ideas with which the researchers attempted to explain the development of the lexeme.

The vowel

Of all the etymologies, only that of Kononov entertains the idea that the original vowel might have been a back vowel. All the others reconstruct a front vowel. The situation is repeated if we consider the labiality of the vowel, with Kononov believing it to be labial, while the rest opt for an illabial vowel. Clauson is alone in reconstructing a long vowel, whereas the rest deem it short. More differences start to arise when considering the height of the vowel. Six etymologies (Radloff, Grønbech, Räsänen, Clauson, Kononov and Hamilton) unambiguously have a broad vowel. M. Stachowski also essentially reconstructs a broad vowel, as in his system there was only one PT *E* type vowel, and whatever the exact pronunciation might have been, it was phonologically broad, as opposed to the narrow **i*. Hartmann employs an inconsistent notation, which makes it impossible to determine whether

he reconstructs a broad *ä* or a narrow *e*. Polivanov argues in favour of an original narrow **e*, in which he is supported by Ramstedt. Bang proposes two alternating forms, one with a broad **ä* and the other with a narrow **i*. In fact, this is the only time that **i* appears in a Proto-Turkic form of the word.

The consonant

There is a universal agreement among the etymologists that the original consonant was labial, but its exact features are disputed. In the opinion of six researchers (Grønbech, Polivanov, Bang, Ramstedt, Räsänen and M. Stachowski) it was the bilabial voiced plosive *b*, although Radloff opts for its voiceless counterpart *p*. Hamilton and Hartmann reconstruct a bilabial voiced fricative *β*, and despite Clauson also viewing the original consonant as a fricative, he deems it the voiced labio-dental fricative *v*. Finally, Kononov reconstructs two alternating forms, one with the bilabial plosive *b*, and the other with the labio-dental fricative *v*.

The ideas utilized in the development models

Some etymologists fail to support their reconstructions with specific argumentation, either limiting themselves to asserting a development chain without any comments, or alternatively, simply naming the phonetic processes. These chains are also not grounded in the actual languages, and there is doubt as whether the author believed that the form found at the end of his chain must have necessarily gone through all stages. Similarly, these chains fail to account for the possibility that two forms that appear identical in different languages might have evolved via different pathways.

It seems that most of the researchers believe that the word under consideration exhibits an irregular development, as they do not refer to any regular sound laws, but only Clauson states this explicitly. Grønbech believes it to be regular, and although Räsänen most likely shares the same sentiment, his view is unclear. Both adduce a few words that developed similarly, among them the verb **säb-* 'to love'. Bang and Kononov compare the lexeme to **sub* 'water', whereas Hamilton and M. Stachowski equate it to **sab* 'word', and its derivatives. Polivanov and Räsänen highlight an analogy with the German *eu > oi* change, although they differ in the specifics. Tatarincev alone discusses semantics. His idea is to derive the meaning 'house' from 'round, curved'. Ramstedt believes the forms *öγ* to be a Mongolian borrowing, while Räsänen considers the forms *uγ* etc. to continue a different, although native, root.

A summary of which researcher proposed which protoform is given in Table 1.

Author	Reconstruction
Radloff (1882)	*äp*
Grønbech (1902)	*äb*
Hartmann (1904)	*eβ*

Author	Reconstruction
Polivanov (1934)	*eb*
Bang (1936)	*äb ~ ib*
Ramstedt (1949)	*eb*
Räsänen (1949)	*äb*
Clauson (1957)	*ā̆v*
Kononov (1958)	*ob ~ ov*
Hamilton (1974)	*äβ*
Tatarincev (2008)	?
Stachowski (2019)	*eb*[14]

Table. 1. Reconstructions proposed by different authors

2. Contemporary data

2.1. Listing of contemporary forms

Below is an alphabetical enumeration of all the forms of the lexeme in question which I was able to find in the modern languages. They are given alongside the languages in which they are present and the sources from which they were taken.

av: Urum (Caucasian) (Verhoeven et al. 2011: 7–11)
äb: Altai, Kondoma, Kyzyl (ÈSTJa I 513) Koibal, Sagai (R I 925)
äm: Shor (Kurpeško-Tannagaševa and Apońkin 1993: 89)
äp: Küerik (R I 916), Middle Chulym (Ölmez et al. 2008: 122; Lemskaja and Kondi-jakov 2021: 53),
äv: Crimean Tatar (Zatov 1906: 26; Useinov 2007: 74), Gagauz (Gucu and Vasiloglu 1996: 28), Karaim (E dialect) (Baskakov et al. 1974: 652), Turkish (TDK), Urum (Caucasian) (Skopeteas et al. 2011: 47), Urum (Ukrainian) (Garkaveć 2000: 610)
äw: Siberian Tatar[15] (Tobol-Irtysh dialect) (R I 671)
čö: Western Yugur (Roos 2000: 401)
ǯü: Western Yugur (Malov 1957: 31; Roos 2000: 401)
ev: Azeri (Širaliev and Oružov 1951: 79)
êp:[16] Koibal (ÈSTJa I 513)

---

[14]   *e = e ~ ä*, see section 1.12

[15]   Radloff uses the designation "Dialekt von Tara". I suspect that this should be interpreted as the Tara dialect of Siberian Tatar, which is classified as part of the Tobol-Irtysh dialect (Tu-maševa 1992: 3).

[16]   It is unclear what this sign is meant to represent. It is missing from the transcription section of the ÈSTJa. A passage below the list of forms in which this form is found suggests treating it as a long *ē*, but this is only conjecture.

ib: Kacha (ÈSTJa I 513) Khakas (Čankov 1961: 193)

ibe: Fu-yü Girgis (Tenišev 1966: 91, 96)[17]

ijö: Western Yugur (Roos 2000: 401)

iv: Karaim (SW dialect) (Mardkowicz 1933: 30; Baskakov et al. 1974: 192)

jiv: Karaim (SW dialect) (Baskakov et al. 1974: 244)

ju: Western Yugur (Malov 1957: 43; Roos 2000: 401)

juv': Karaim (NW dialect) (Baskakov et al. 1974: 254)

jö: Western Yugur (Roos 2000: 401)

jü: Western Yugur (Malov 1957: 45; Roos 2000: 401)

jüä: Western Yugur (Malov 1957: 45; Roos 2000: 401)

jüj: Western Yugur (Roos 2000: 401)

jüö: Western Yugur (Roos 2000: 401)

jüv: Karaim (NW dialect) (Baskakov et al. 1974: 260)

oj: Uzbek dialects (ÈSTJa I 513)

öγ: Karagas (ÈSTJa I 513), Soyot (Rassadin 2006: 103), Tofalar (Rassadin 2005: 176), Tuvan (Tenišev 1968: 330)

öγ': Tuvan (ÈSTJa I 513)

öj: Bashkir (Uraksin 2005: 292), Siberian Tatar (Tobol-Irtysh dialect) (Sagidullin 2010: 43), Taranchi (R I 1171), Tatar (Ganiev 1997: 135), Turkmen (Baysähedow 2009: 74), Uyghur (Necip 1995: 309), Lobnor, Uzbek dialects (ÈSTJa I 513)

öm: Shor (Matyr dialect/Mrass dialect) (?)[18] (R I 1313)

öp: Sagai (R I 1308)

öv: Azeri (R I 1313)

śü: Western Yugur (Roos 2000: 401)

ū: Lebedin (Baskakov 1985: 215), Shor (R I 1591)

uγ: Lebedin (Baskakov 1985: 215)

ūγ: Shor (R I 1617)

uj: Uyghur (ÈSTJa I 513), Uzbek (Awde et al. 2002: 73)

uw: Altai dialects (ÈSTJa I 513)

ü: Baraba Tatar (Dmitrieva 1981: 193), Tatar dialects (ÈSTJa I 513), Western Yugur (Malov 1957: 132)

ū: Baraba Tatar, Shor, Simbir[19] (R I 1797) Western Yugur (Roos 2000: 401)

üγ: Lebedin (Baskakov 1985: 215), Lower Chulym (Dul'zon 1973: 20), Shor, Tuvan (R I 1807)

---

[17] I would like to thank the anonymous reviewer for directing me to this source.

[18] Radloff's description of the source of this form reads "Mad. W.", but the abbreviation is not explained in his work. It is not even entirely certain whether it refers to a language or a written source. The interpretation of this reference as "Matyr" is based on the opinion of K. Stachowski (2021: 39). The anonymous reviewer has in turn suggested reading this source as Mrass. However, determining the exact dialectal source of this form is not central to our study.

[19] It is uncertain where this dialect should be located. It is perhaps connected to the city of Ulyanovsk, which until 1924 bore the name of Simbirsk. In that case, the "Simbirsk dialect" should most likely be understood as a dialect of Tatar.

üj: Altai (Baskakov 1964: 142), Baraba Tatar (Dmitrieva 1981: 193), Karachay-balkar (Chabičev 1990: 70), Karaim (Crimean dialect) (Baskakov et al. 1974: 589), Karakalpak (Baskakov 1967: 201), Kazakh (Bektaev 2007: 537), Kondoma (Kurpeško-Tannagaševa and Apońkin 1993: 89), Kumyk (KS), Kyrgyz (Judachin 2000: 171), Lebedin (Baskakov 1985: 215), Nogai (Baskakov 1956: 161–162), Teleut (Rjumina-Syrkaševa and Kučigaševa 1995: 94), Uzbek dialects (ÈSTJa I 513), Western Yugur (Malov 1957: 132; Roos 2000: 401)

üje: Western Yugur (Roos 2000: 401)

üv': Karaim (NW dialect) (Baskakov et al. 1974: 588)

üw: Altai dialects (ÈSTJa I 513)

üχ: Soyot (ÈSTJa I 513)

žüä: Western Yugur (Roos 2000: 401)

As can be seen, the material is wide-ranging and comprehensive. Attempting to directly compare all the existing forms would not only be impractical, but the result of such an operation would most likely be chaotic and difficult to digest. However, selectively ignoring certain forms could lead to serious mistakes, with the criteria determining which forms should be retained and which ignored are by no means clear-cut. This problem would be solved if certain forms were grouped together and then compared, with the most straightforward approach focusing on language groups such as Oghuz, Kipchak, etc. However, when considering the available data it becomes clear that the variety of forms does not align with the usual classifications. One language can have many different forms and it is not always simple to assign these forms to particular dialects, which is why a "reduction of forms" has to be based on the forms of the lexeme, and not on the languages. However, it would be somewhat surprising if every single form from the list above developed directly from Proto-Turkic and independently from other languages. There must have been some intermediary forms, from which the modern forms emerged. If these intermediary forms can be established, this would provide natural groupings, which would provide the ideal material for further comparison. Determining these intermediaries, which in this study will be called "pre-forms" (as in Pre-Karaim, Pre-Altai, etc.), will be the aim of the present section. The purpose of this operation is strictly utilitarian, as we want to go as far back as the available information will allow. As the state of research on individual languages differs, the result is likely to be asynchronic, but this should not have any significant impact on the results. Therefore, each language will be addressed, in an attempt to establish a pre-form for each.

The method employed for this analysis will be based almost exclusively on the application of the established sound laws. In the event of a lack of appropriate sound laws, original observations on the sound systems of the discussed languages will be made. Comparison with other languages will be used sparingly in the reconstruction of the pre-forms, since the current procedure is intended to prepare the data for the subsequent comparisons and reconstructions. As the state of the Turkic historical and comparative grammar is still unsatisfactory, it will not be possible

to trace back some forms further than their contemporary shape. If no appropriate sound laws exist which could explain the forms found in a certain language, this language will not be discussed in this section. In such cases, the modern forms should be viewed as equivalent to the pre-forms. The results of the analysis will be given in section 2.17.

For ease of navigation, the section is arranged according to the usual classifications of the Turkic languages. The groups will be discussed in the following order: Oghuz, Karakhanid, Kipchak, South Siberian. In each of the groups, the languages will be arranged alphabetically.

2.2. The Azeri forms

ev (Širaliev and Oružov 1951: 79)
öv (R I 1313)

Azeri differentiates between a narrow *e* and a broad *ä* (Širaliev and Sevortjan 1971: 6–12). This distinction has been used in the past as one of the arguments to propose a similar distinction for Proto-Turkic (Doerfer 1975–1976b: 14–15). It has subsequently been critiqued, and the issue has not yet been satisfactorily resolved (M. Stachowski 1998: 162–164). Considering the current situation, the Azeri *ev* can be derived from either **ev* or **äv*. As the former would be a completely isolated form, the latter, which is identical to the shape commonly found in other Oghuz languages, will be accepted as the Pre-Azeri form. The issue of the narrow *e* in the reconstruction of this lexeme will be discussed in more detail in section 4.2.

The labialization of low vowels before labial consonants is a known phenomenon in Azeri (Ščerbak 1970: 40). Thus, in all likelihood, *öv* also goes back to the Pre-Azeri **ev ~ *äv*.

2.3. The Caucasian Urum forms

äv (Skopeteas et al. 2011: 47)
av (Verhoeven et al. 2011: 7–11)

The form *äv* is the typical Oghuz shape of the lexeme in question. No sound laws which could have produced it on Urum grounds are known to me, which is why this form will be treated as primary.

It is uncertain whether the form *av* represents real data or is a result of a typing mistake. It is found solely in the Caucasian Urum language, and moreover in a single source, while the only other source offers *äv*. This fact raises doubts, as the data for both of these sources was gathered in the same area and during the same time period (Tbilisi, October – November 2010 (Skopeteas et al. 2011: 10; Verhoeven et al. 2011: 5)). Thus, we would expect the data from the two sources to be identical, but unfortunately the authors fail to provide any explanation for the discrepancies. This seems to suggest that *av* is simply an erroneous notation. However, even if

this conjecture is proven wrong and *av* is a real form, it would still most likely be a secondary development which would go back to an earlier **äv*.

As it stands, both of the attested forms most likely go back to **äv*, which should be considered the Pre-Urum form.

2.4. The Turkmen form

öj (Baysähedow 2009: 74)

As Turkmen is generally considered an Oghuz language, one might expect that this form would go back to the common form found in other Oghuz languages, namely *äv*. Indeed, the Turkmen *e* before *w* does undergo labialization (Clark 1998: 29–31), but this process never results in a consonant change. Thus, *öj* cannot be derived directly from **ew* (< **äv*). As there are also no other known sound laws which could explain this form, it will be treated as the Pre-Turkmen form, distinct from the Oghuz *äv*.

2.5. The Uyghur forms

öj Taranchi (R I 1171), Uyghur (Necip 1995: 309), Lobnor (ÈSTJa I 513)
uj Uyghur (ÈSTJa I 513)

Regarding *öj*, the situation is similar to that in Turkmen, with a tendency towards the labialization of the vowel before a labial consonant (Ščerbak 1970: 40). However, it does not lead to a simultaneous consonant change, and because of this, the direct development of *öj* from **äv* is impossible, and *öj* has to be regarded as the Pre-Uyghur form.

There are no phonological processes that could regularly produce the form *uj*. The so-called "Uyghur umlauting" cannot be called upon in this instance, as firstly, it only raises low unrounded vowels into a narrow *e*, and secondly it does not explain the backing of the vowel (Yusup and Memet 2013). This form does not pertain to the literary language, so a likely explanation is that it comes from a dialect influenced by the Uzbek language. This interpretation will be accepted in this study, and thus the form will not be treated as primary.

2.6. The Uzbek forms

öj (ÈSTJa I 513)
oj (ÈSTJa I 513)
üj (ÈSTJa I 513)
uj (Awde et al. 2002: 73)

The first thing to note is that only the final form comes from the literary language, while the remainder are assigned the nomination "Uzbek dialects" in the ÈSTJa. The classification of Uzbek dialects is a complex issue. A brief, yet useful, summary

can be found at CTILD. Among other things, one of the features that are used to distinguish the Uzbek dialects is the retention or loss (under Iranian influence) of the distinction between front and back rounded vowels (cf. Polivanov 1929; Rešetov 1957: 5–30). We know that the literary language belongs to the group which lost this distinction (Sjoberg 1963: 3), and it is known with considerable certainty that *oj* can also be ascribed to such dialects. As an example, the Tashkent dialect, which is also an Iranized dialect, contains exactly this form (Guljamov 1957: 168). Thus, the emergence of *oj* and *uj* can be explained by the merging of the rounded back and front vowels. Accordingly, the palatal shapes must be assigned to the dialects that retained the opposition of the two sets of rounded vowels. Consequently, all the Uzbek forms can be traced back to *öj* and *üj*. As for the height difference, we can only note its existence, as the available sources do not allow us to unambiguously establish the relations between them. According to Kononov (1960: 19), the modern Uzbek *u* goes back exclusively to *u* and *ü*. No mention is made of any process that could have otherwise produced *u*. As for the origin of *o*, he only writes that it is one of the reflexes of an earlier *a*, which does not assist in establishing a relation between *öj* and *üj*. A relevant question is whether such a relation even exists, as different dialectal forms do not have to go back to the same "Proto-Uzbek" form as the literary form, since there was no such thing as Proto-Uzbek (Polivanov 1933: 4).

2.7. The Western Yugur Forms

üj (Malov 1957: 132)
üjä (Roos 2000: 401)
ju (Malov 1957: 43)
jü (Malov 1957: 45)
jüj (Roos 2000: 401)
jüä (Malov 1957: 45)
jüö (Roos 2000: 401)
jö (Roos 2000: 401)
ijö (Roos 2000: 401)
ǯü (Malov 1957: 31)
ǯüä (Roos 2000: 401)
śü (Roos 2000: 401)
čö (Roos 2000: 401)
ü (Malov 1957: 132)
ǖ (Roos 2000: 401)

Western Yugur is not a well-studied language, and the available grammatical descriptions often contradict each other. In addition, the variety of forms found in WY is unmatched in the other Turkic languages, a fact which can be ascribed to the lack of an established norm of the language and the origin of the data, which come from various field work expeditions. Because of this, not all the forms can be convincingly explained, but some hypotheses about the Pre-Western Yugur form can be made.

Some forms can be ascribed to an erroneous notation. The *ǖ* is recognized by Roos as coming from Malov's work, but this is uncertain, as in the edition of Malov's work that I examined, a certain sign above *ü* could be interpreted as a length marker, but it could also be a printing mistake (Malov 1957: 132). Perhaps the copy Roos used contained a similar mark, which influenced his interpretation. Roos (2000: 15) *explicite* rejects the existence of long vowels in WY, and Malov fails to mention vowel length at all. Consequently, I decided to read Malov's entry as a short *ü*. As for *žüä*, it is again attributed by Roos to Malov. Indeed, such a form is found as a cross-reference under the heading *jü*, but the entry to which it refers is missing (Malov 1957: 45). The form is suspicious, as there are a total of four entries with initial *ž-* cited by Malov (1957: 31).

All the forms with an initial consonant other than *j-* are problematic. Roos (2000: 27) mentions that one of the allophones of /j/ before *ü* is *ź*, which could somehow be connected with *śü*. Malov (1957: 159) states that *ǯ* is the result of the evolution of *j* in front of *i* and *y*, but the relevant section in his dictionary reveals many more words with an initial *ǯ-* followed by different vowels, among which *i* and *y* are not exceptionally numerous (Malov 1957: 28–31). The available reference works do not provide any information which could help explain the emergence of the forms *žüä* and *čö*. Despite this, the assumption that all these shapes stem from an earlier **j-* seems reasonable.

As for the forms with more than one vocalic element, the sources fail to describe any phonetic processes that could have produced them. Perhaps they can be explained as suffixed forms, with the suffix possibly being the possessive suffix, although such a solution is also debatable, as the expected suffix should have been *-si* (Roos 2000: 68). In all likelihood, these forms are secondary and can be traced back to forms with a single vowel.

This leaves us with the forms *ju*, *jü*, *jüj*, *jö*, *üj*, *ü*. Two models can be proposed to explain their origin. Roos (2000: 52) believes the development chain is as follows:

CT **äb > *ö > jö > jü*

To account for the other forms, I believe that this model can be expanded in the following way:

*ö > 1) jö > jü; 2) ü > *u > ju

The origin of **ö* has been ignored in this scheme, as it is not strictly a WY issue, but pertains to the broader comparative picture, which will be addressed in section 4.2. Roos also does not discuss it further.

In this model, the form *ü* came to be as a result of vowel raising, which is a regular phenomenon (Tenišev and Todaeva 1966: 11–12; Roos 2000: 13, 42–43), and the same change has to be assumed for the *jö > jü* development. The emergence of a prothetic *j* is a known, although irregular phenomenon (Roos 2000: 41, 43–44). The form *ju* is problematic. The development **ü > u* is regular in WY, but it did not occur after *j* (Roos 2000: 39). Therefore, *ju* must be derived from *ü > *u*, and not from *jü*. Despite it, this development is still unexpected, as *j-* typically not

only stopped the velarization of the vowel, but also fronted originally back vowels (Roos 2000: 39).

As can be seen, this model has certain drawbacks. In addition to those already discussed, the chronology of the vowel raising is not clear. In the historical section, Roos treats this process as if it has already been completed (Roos 2000: 42–43). However, when reporting the issues encountered during her research (Roos 2000: 11), and in the descriptive section (Roos 2000: 12–13) she states that it is a very recent, still unfinished trend. If the latter opinion is correct, then it becomes a serious obstacle for the proposed model. It would be expected that the older forms with a broad vowel would be more numerous in the data and that they would be more prominent in the older sources, while giving way to the new forms in more recent surveys, but this is not the case. Another problem is that this model fails to account for *üj* (and, consequently, for *jüj*, which most likely should be derived from the former).

In view of these difficulties, let us now consider an alternative model.

üj > 1) *jö*; 2) *jüj*; 3) *uj* > *ju*; 4) *jü* > *ü*.

The form *jüj* came to be through the addition of a prothetic *j*, with the forms *ju*, *jö* and *jü* the result of a metathesis (Malov 1957: 132, 162; Ayazlı 2019: 21). However, the back vowel in *ju* is just as problematic in this model as in the previous one, and the presented solution is similarly unconvincing. The lowering of the vowel is not regular, but other examples of a similar development are known (Roos 2000: 53). Finally, the form *ü* developed from *jü* through apheresis, which is known to have occurred frequently in WY (Roos 2000: 41, 43).

In summary, there are numerous problems and question marks surrounding the development of the abundant variants of the word in question in Western Yugur. Neither of the proposed models is ideal, although I believe the second to be more plausible, as it faces fewer obstacles. This means that the Pre-Western Yugur form should be reconstructed as *üj*. All the other forms can be derived from it, albeit with varying degrees of certainty.

2.8. The Bashkir-Tatar form

öj Bashkir (Uraksin 2005: 292), Tatar (Ganiev 1997: 135),

The form found in these languages can be explained by the Volga Vowel Shift, which is a process which raised the PT low labial vowels *ö* and *o* to *ü* and *u*, and, simultaneously, lowered the PT high labial vowels *ü*, *u* to *ö* and *o* (Tekin 1987). Thus, the form *öj* in these two languages has to be derived from an earlier *üj*.

2.9. The Karaim forms

iv SWK (Mardkowicz 1933: 30; Baskakov et al. 1974: 192)
jiv SWK (Mardkowicz 1933: 36; Baskakov et al. 1974: 244)

juv′ NWK (Baskakov et al. 1974: 254)
jüv NWK (Baskakov et al. 1974: 260)
üv′ NWK (Baskakov et al. 1974: 588)
äv EK (Baskakov et al. 1974: 652)
üj EK (Baskakov et al. 1974: 589)

The vocalism of the SW forms is the result of the change *$*ü > i$* (Baskakov et al. 1974: 11). As for the initial *j-*, Mardkowicz (1933: 3–4) states, that "[t]he spelling of some Karaim words beginning with *i*, *j* or *y* is not strictly established".[20] He does not comment on the reason for this alternation, so it is uncertain whether such notation reflects the real pronunciation or is simply a graphical convention. In any case, both forms have to be derived from *$*üv$*.

All the NW forms are the result of the harmony shift from vowel to consonant harmony in NW Karaim. There is extensive literature concerning this phenomenon, for a summary of which, cf. K. Stachowski (2009). The basic premise of this change is that the vowels lost their frontness, which was shifted to the consonants in the form of palatalization (as in the Slavic languages). Nevertheless, these forms are somewhat peculiar, as there appears to be confusion among the specialists regarding this harmonic shift in the absolute anlaut. On the one hand, Németh (2014: 354) states: "There was no *ö-* > *′o-* and *ü-* > *′u-* change in the initial position", yet on the other, he also cites Pritsak (1959: 327) and Hamp (1976: 212), and ascribes to them the opinion that: "word-initially, *ü-* may have alternated with *ju-*." (Németh 2014: 354). As our example shows, the latter point of view is without doubt correct. There remains, however, a second problem, namely the existence of three alternating forms. To explain this, it will be useful to mention the observation of Musaev (1964: 49): "In the speech of the young generation, *ü* in almost all positions of the word (among them in the anlaut) loses its frontness and is pronounced like the diphthong *ju*".[21] In light of this, we can assume that these forms constitute three chronological stages of the development of the harmony shift in the anlaut. The form *üv′* is the oldest, *jüv* marks the transition point, and *juv'* is the shape expected after the shift has been completed. The only minor issue that remains is the absence of the palatalization of the final *v* in *jüv*, but this can reasonably be ascribed to the unusual pronunciation of certain individuals. All these forms must originate from an earlier *$*üv$*.

As for the Eastern Karaim forms, I know of no sound laws that could explain them. As a result, they cannot be traced further back on Karaim grounds. However, the presence of a pair of forms that are not directly related can be explained. The form *äv* is characteristic of the Oghuz languages; in Karaim it is most likely a borrowing from Crimean Tatar.[22] The form *üj* is the shape found in almost all

[20] "Pisownia niektórych wyrazów karaimskich, mających początkową literę i, j lub y nie jest ściśle ustalona."

[21] "В речи молодого поколения ÿ почти во всех позициях в слове (в том числе и в анлауте) теряет палатальность и произносится как дифтонг йу."

[22] While Crimean Tatar is a Kipchak language, it having been strongly influenced by Turkish is well known.

the Kipchak languages (interestingly, the exception being the western dialects of Karaim).

In summary, Eastern Karaim exhibits two form, *üj* and *ev*. The first shows a typically Kipchak development, while the second is a borrowing from Crimean Tatar. The Western Karaim forms have a different origin and can all be traced back to *üv.

2.10. The Siberian Tatar forms

äẅ Siberian Tatar (Tobol-Irtysh dialect) (R I 671)
öj Siberian Tatar (Tobol-Irtysh dialect) (Sagidullin 2010: 43)
üj Baraba Tatar (Dmitrieva 1981: 193)
ü Baraba Tatar (Dmitrieva 1981: 193)
ū Baraba Tatar (R I 1797)

The form *äẅ* (symbolizing a diphthong with a *ü* coloured labial glide (Sagidullin 2008: 13–14)) cannot be traced further back to Siberian Tatar grounds.

As for *öj*, the Tobol-Irtysh dialect of Siberian Tatar, which belongs to the same areal grouping as Tatar and Bashkir, also took part in the Volga Vowel Shift (Tumaševa 1992: 4). Thus, this form must be derived from an earlier *üj.

Monophonemic forms likely go back to a single form. Dmitrieva (1981: 202–203) fails to provide a clear account of vowel length in Baraba Tatar. On the one hand, she states that the phenomenon is not phonemic and appears positionally, although she does not specify the positions. On the other hand, she gives examples of phonemic vowel length emerging from contractions. It is most likely an issue of dialectal diversification, with some dialects showing phonemic vowel length, whereas in others it is an unimportant phonetic feature. When discussing the vowel length resulting from contractions, she states that the loss of a neighbouring consonant can also result in vowel lengthening, which in our case could imply the development *üj* > *ū*. However, the examples adduced by her do not illustrate such a process, being limited to syllable contractions.

This leaves us with three distinct Pre-Siberian Tatar shapes: *äẅ, üj* and *ū*. The last two could possibly be connected, but this is not certain.

2.11. The Altai forms

äb Altai (ÈSTJa I 513)
üw Altai dialects (ÈSTJa I 513)
uw Altai dialects (ÈSTJa I 513)
üj Altai (Baskakov 1964: 142), Lebedin (Baskakov 1985: 215), Teleut (Rjumina-Syrkaševa and Kučigaševa 1995: 94),
ū Lebedin (Baskakov 1985: 215)
uγ Lebedin (Baskakov 1985: 215)
üγ Lebedin (Baskakov 1985: 215)

As I know of no sound laws which could explain the form *äb* on Altai grounds, it will be treated as primary.

The forms *ū*, *uγ* and *üγ*, as well as *üw* and *uw* are likely to be connected. The vowel length in Lebedin is usually the result of the elision of the final *γ* (Baskakov 1985: 14), so *ū* can be traced back to *uγ*. The alternation of the final *-γ* and *-w* is also known (Baskakov 1985: 11, 14), as is the alternation of the high vowels *u:ü* and *y:i* (Baskakov 1985: 18). Although this allows us to connect these five forms into a single pre-form, it does not clarify how that form should be reconstructed, as both **-γ* and **-w* would have yielded the same alternation [cf. *taγ ~ taw ~ tū* 'mountain' < **tag*, and *suγ ~ suw ~ sū* 'water' < **sub* (Baskakov 1985: 11)]. Regarding the vowel, the abundance of front vowel forms and the lack of back vowel forms in other languages does suggest a reconstruction with a front *ü*. With regard to the consonant, I am inclined to prefer the *-γ* form, as it parallels the Shor and Tuvan forms, while the only parallel to the potential **üw* form can be found in Western Karaim. Thus, these five forms can be tentatively derived from **üγ*.

In Lebedin and related dialects, in the intervocalic environment at the morpheme boundary, *γ* has a tendency to change into *j* (Baskakov 1985: 11). Thus, the form *üj* could be explained as secondarily derived from a suffixed form in which this process has taken place, but it is far from certain. Additionally, this process is not mentioned in the available grammars of the literary language, and seemingly contradicts the elision of *γ* and the subsequent lengthening of the vowel (Dyrenkova 1940: 24). Although the existence of both the contracted forms and the forms with *γ > j* are clearly conceivable, as evidenced by the dialectal data discussed above, *üj* could also be a pre-form in its own right, rather than being derivable from *üγ*.

Three Pre-Altai forms can be reconstructed, namely **äb*, **üγ* and *üj*. The last two forms might also be related to each other. The picture is further complicated by the possibility of reconstructing **üw* instead of **üγ*, which raises a question as to how *üj* could fit into that scheme.

2.12. The Chulym forms

äp Küerik, Middle Chulym (Ölmez et al. 2008: 122; Lemskaja and Kondijakov 2021: 53; R I 916)
üγ Lower Chulym (Dul'zon 1973: 20)

Chulym Turkic is severely understudied and sadly no historical phonology of the dialects is available. Different descriptions of Chulym Turkic sometimes contradict each other. An example of this, which is somewhat linked to our study, is the case of the voiced labial phonemes. Pomorska (2001: 89), as well as Lemskaja and Kondijakov (2021: 23), propose the labial voiced fricative *v* as an independent phoneme, while Ölmez et al. (2008: 20–21) treat it as an allophone of /b/.

This is relevant for the form *äp* in Middle Chulym, in which the voiceless *p* in *äp* alternates with *b ~ v* according to the phonotactic rules of the language. While *p* is found in the anlaut, auslaut and consonant clusters with other voiceless consonants,

b ~ v replaces it in the intervocalic position and after voiced consonants (Pomorska 2001: 92–93; Ölmez et al. 2008: 21). This means that the form can be equally well traced back to **äp*, **äb* and **äv*, as all of these would have yielded the same result in the modern language. That said, comparative data do favour the voiced auslaut, so the reconstruction of Pre-Chulym **äb* will be accepted here.

The situation is similar with the Lower Chulym dialect, which is even less researched, and the data concerning it are scarce. For these reasons, the modern form *üγ* will have to suffice and be used for comparison with the other languages.

2.13. The Khakas forms

äb Koibal, Sagai (R I 925), Kyzyl (ÈSTJa I 513)
êp Koibal (ÈSTJa I 513)
ib Khakas (Čankov 1961: 193), Kacha (ÈSTJa I 513)
öp Sagai (R I 1308)

As already highlighted in footnote 16, it is not clear how the vowel in *êp* should be read. If it truly is a long vowel, as Sevortjan's wording seems to suggest,[23] its presence is difficult to explain. Long vowels in Khakas are the result of contractions, but the lexeme under consideration does not display polysyllabic variants that could have been contracted.

The form *ib*, found in the literary language, shows the characteristic narrowing of the vowel *ä* in the initial syllables. Although this process is poorly documented in the available Khakas grammars, its existence is evident from such forms as: *it* 'meat' (Subrakova 2006: 138), *kil-* 'to come' (Subrakova 2006: 162), *kirek* 'required' (Subrakova 2006: 169), *pis* 'five' (Subrakova 2006: 362).

According to Baskakov (1975: 38) *b* cannot appear in the auslaut in the literary language. As the adduced examples clearly show, this rule is not reflected in the orthography of the literary language. It is responsible for the devoicing of the final consonants in *êp* and *öp*.

The possibility of labialization of the vowel by a following labial consonant as in *öp* is not asserted by the grammarians, but nonetheless it can be assumed to be the most likely explanation of this form.

Although not all the intricacies of the Khakas forms have been solved, it is probable each of them can be derived from the Pre-Khakas **äb*.

2.14. The Fu-Yü Girgis forms

ibe (Tenišev 1966: 91, 96)

In Fu-Yü Girgis, similarly to Khakas, **ä* in initial syllables undergoes raising into *i* (Tenišev 1966: 91). The origin of the final *e* is uncertain, but in all likelihood it is secondary. Thus, the pre-form can be reconstructed as **äb*.

---

[23] "The length of the root vowel in Koibal has no parallels in other languages (Долгота корневого гласного в койбальском не имеет параллелей в других языках)" (ÈSTJa I 513).

2.15. The Shor forms

äm (Kurpeško-Tannagaševa and Apońkin 1993: 89)
öm Matyr/Mrass (?) (R I 1313)
äb Kondoma (ÈSTJa I 513)
üj Kondoma (Kurpeško-Tannagaševa and Apońkin 1993: 89)
ū Shor (R I 1591)
ǖ Shor (R I 1797)
ūγ Shor (R I 1617)
üγ Shor (R I 1807)

The form *äm* can be explained by the alternation of the labial consonants in different dialects; where one dialect has *b*, the other displays *m* (Dyrenkova 1941: 19). The literary language retains many words with *m* which correspond to *b* in other dialects and languages. As *m* is limited to Shor, this allows us to derive *äm* from an earlier *äb*. As the identification of the source of the form *öm* is tentative (see footnote 18), nothing certain can be said about its origin and development. It is most likely also derivable from *äm* and further from *äb*.

Regarding Radloff's forms, those without a consonant can in all likelihood be understood as a result of an elision of the auslaut *γ*. This also explains the length of the vowels, as in Shor the vowel length is the result of such elisions (Dyrenkova 1941: 10–11). Thus, the number of shapes can be reduced to two: *ūγ* and *üγ*. The first form carries the vowel length even though the consonant is intact, which is difficult to explain. Perhaps it was an intermediate state before the vowel was completely lost, or perhaps it is simply an unfortunate mistake by Radloff.

The back vowel forms are difficult to explain, but in any case it is not very likely that they present archaic features of significance to the reconstruction. A more detailed discussion of these forms is to be found in section 4.1.2.

The origin of *üj* may also be related to *üγ*. A tendency towards the *γ > j* change is known in some dialects, although it usually only happens intervocalically (Dyrenkova 1941: 17). Perhaps this change occurred in a possessive form and the unsuffixed form was subsequently changed by analogy, or perhaps there are instances where this process can also take place word-finally. An alternative solution would be to not connect these forms directly, and instead to relate *üj* to the multitude of languages where it is primary.

In summary, all the shapes listed above can be derived from three distinct Pre-Shor forms, **äb*, **üγ* and *üj*. There is a possibility that the latter two also go back to a single form, but it is not certain.

2.16. The Tuvan forms

öγ Tuvan (Tenišev 1968: 330), Soyot (Rassadin 2006: 103), Tofalar (Rassadin 2005: 176), Karagas (ÈSTJa I 513)
öγ′ Tuvan (ÈSTJa I 513)

üγ Tuvan (R I 1807)
üχ Soyot (ÈSTJa I 513)

The forms *öγ* and *öγ′* should be understood as being identical. The palatalization of *γ* is due to its position in a front vocalic word and is not phonemic. Being allophonic, it is usually not marked in writing, making Sevortjan's notation unique (Ischakov and Pal'mbach 1961: 53–54).

Although the grammatical descriptions of Soyot do not account for the word-final devoicing of *γ*, such a process should probably be assumed here, linking *üγ* and *üχ* together.

This analysis allows us to reduce the number of Tuvan forms to two, *öγ* and *üγ*. While it does seem to be apparent that they are related, their relationship is unclear. There is no regular vowel raising or lowering in the languages in question. On the one hand, the form with the low vowel is much better attested, with citations coming from modern dictionaries. On the other hand, the high vowel form has parallels in neighbouring languages beyond the Tuvan dialects. The low vowel form had been interpreted by some as a Mongolian loan (cf. Ramstedt 1949: 56), which could suggest that these forms are of a different origin, but this hypothesis has also been rejected by some (Tatarincev IV 354–355). This will be discussed in more detail in section 4.1.1. For now, it seems impossible to derive one of these forms from the other, and so they will both be treated as primary.

2.17. The reconstructed pre-forms

The analysis of the pre-forms allows us to establish nine groups. However, the groups **äẅ* and **ü* seem to be the result of a certain degree of ignorance more than anything else. Although our current knowledge of Siberian Tatar dialects does not allow us to reliably assign them to another group, it is extremely unlikely that they truly represent a completely separate development from Proto-Turkic. Further studies should be able to link these forms to other groups, but in this study these two groups will be excluded from the analysis, while the remaining seven will be compared in section 4.2 The geographical distribution of these groups is presented in Map 1.

*äb: Shor, Sagai, Kondoma, Altai, Kyzyl, Middle Chulym, Küerik, Koibal, Khakas, Kacha, Fu-Yü Girgis
*äv: Azeri, Gagauz, Turkish, Urum (Caucasian), Karaim (E dialect), Crimean Tatar, Urum (Ukrainian)
*äẅ: Siberian Tatar (Tobol-Irtysh dialect)
*öγ: Tuvan, Soyot, Tofalar, Karagas
*öj: Turkmen Taranchi, Uyghur, Uzbek dialects, Lobnor
*ü: Baraba Tatar
*üγ: Altai dialects, Shor, Tuvan, Lebedin, Lower Chulym, Soyot
*üj: Baraba Tatar, Bashkir, Karachay-Balkar, Karaim (E dialect), Karakalpak, Kazakh, Kyrgyz, Kumyk, Nogai, Altai, Kondoma, Lebedin, Tatar, Teleut, Uzbek dialects, Western Yugur, Uzbek
*üv: Karaim (SW dialect), Karaim (NW dialect)
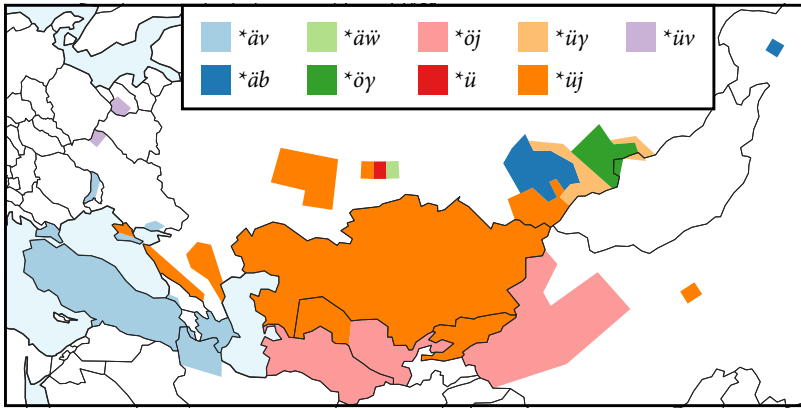
Map 1. The Areal distribution of the pre-form based groups.

3. Historical data

3.1. Listing of the proposed readings of the historical forms

äv: Oghuz (MK I 85 (27)), Kipchak (Grønbech 1942: 96; Toparlı et al. 2007: 77)
ā̆v: Orkhon (EDT 3–4)
äw: Karakhanid (MK I 85 (27))
eb: Orkhon (DTS 162)
ef: Karakhanid (DTS 165)
iv: Kipchak (Grønbech 1942: 96; Toparlı et al. 2007: 117)
jäv: Kipchak (Toparlı et al. 2007: 320)
ov: Armeno-Kipchak (Tryjarski 1969: 595)
öj: Kipchak (Toparlı et al. 2007: 213)
öv: Armeno-Kipchak (Garkavec 2010: 1103–1105), Kipchak (Grønbech 1942: 96; Toparlı et al. 2007: 213)
üj: The Legend of Oghuz Khan dialect[24] (DTS 623)
üv: Karakhanid (DTS 629), Armeno-Kipchak (Garkavec 2010: 1103–1105)

The considerable variety of forms is partly due to the incompatible transcription conventions used by different authors, rather than to real material differences. Therefore, it would be useful to reflect upon these discrepancies in order to achieve a better correspondence with the actual data. This is done below and, additionally, some unusual forms will also be discussed.

3.1. The *j-* form

The form *jäv* is unusual, and although it is probably an innovation going back to *äv*, its emergence is not clear. Similar forms are found in modern Western Karaim

---

[24] The language of this work is uncertain.

dialects, but those are the result of a later development, so this form cannot be linked to them directly. It is tempting to interpret it as a trace of a long *$\bar{a}$ in the etymon[25] (Tekin 1994, 1995: 138–145), but this is unlikely. Firstly, this is an isolated form, found in a single historical source. Were this really a prothesis resulting from the original vowel length, this form would be expected to also be found in Chuvash and some of the Karakhanid languages (not Western Yugur, as the protheses found in this language are of a different origin, cf. section 2.7 and Tekin 1994: 61). Secondly, all the other signs point towards a reconstruction with a short vowel. Thus, this hypothesis cannot be accepted. Tekin also mentions the possibility of a prothetic *j-* corresponding to the Haladj *h-*, but he himself limits this to the vowels *i*, *y* and *ü*, so this explanation does not apply in this case (Tekin 1994: 52). Therefore, the origin of this form requires further study.

3.2. The *ä-*, *e-* forms

As to the form *ā̆v*, Clauson consistently writes a long *ā* in forms attested in the runic script, despite the fact that this writing system did not distinguish vowel length, at least not in a consistent manner (Róna Tas 1991: 58–59; Erdal 2004: 47–48).Thus, this form should be equated with *äv*.

The form *ef* is cited by DTS based on Maḥmūd al-Kāšgarī's work, but it is unfortunately most likely an erroneous reading. Dankoff and Kelly's edition does not feature *ef*. The *-f* in this form is most likely a rendering of the so-called "thin *fā*". Al-Kāšgarī himself explained that he used this letter for a sound between *f* and *b* (MK I 55). Dankoff and Kelly transcribe it with *w* (MK I 55), and argue that it might have been either a glide [w] or a bilabial fricative [β]. It is hard to accept that this sign should be read as [f].

Both this form, and the form *eb* are cited by DTS with a narrow *e*. Dankoff and Kelly consistently write Maḥmūd al-Kāšgarī's form with a broad *ä*, as does Clauson for the runic form. DTS does distinguish between two types of *E*, but their distribution is striking. There are much fewer words with *ä* in this dictionary than there are with *e*. Additionally, almost all the lexemes with *ä* are evident borrowings. While the authors do not explicitly discuss this issue, it seems they believe that *ä* was not a native phoneme in Old Turkic. As such, their notation fails to be consistent with the notations of other scholars, who presume a phonological opposition between /e/ and /ä/ in native words. Because of this, I believe the forms *ef* and *eb* should not be contrasted with *äw* and *ā̆v*, but are to be taken as different readings of the same attestations. The difference in consonants between *ā̆v* and *eb* stems from different interpretations of the runic script, with the sign used in this word traditionally rendered as *b* in all positions. However, some scholars, among them Clauson (1962: 77), believed that post-vocalically it should be read as *v*. For a more thorough discussion of the issue see Erdal (2004: 63–67) and section 4.2.

---

[25] I would like to thank the anonymous reviewer for notifying me of this possibility.

3.3. The forms with a labial vowel

The form *üv* is ascribed by DTS to Maḥmūd al-Kāšġarī and Qutadgu Bilig. However, the former citation is doubtful. Dankoff's and Kelly's edition does not contain this form, nor anything that would resemble it. Perhaps citing Maḥmūd al-Kāšġarī in this instance is simply an unfortunate error on the part of the authors of the DTS, and the form pertains solely to Qutadgu Bilig.

As far as Armeno-Kipchak forms are concerned, there is a difference of opinion regarding the transcription of the Armenian script. Tryjarski (2017: 14–15, 35) believes that the Armenian alphabet does not differentiate between front and back labial vowels, whereas Garkavec (2010: 9), on the other hand, maintains that the front labial vowels can be distinguished in the script. Regrettably, he fails to explain his methods. The result of this discrepancy is that Tryjarski transcribes every labial vowel as back, while Garkavec does not. This is the source of the difference between the forms *öv* and *ov*. While I am inclined to agree with Tryjarski that the script is ambiguous in this regard, it does not mean that the language itself did not distinguish between front and back labial vowels. However, even if it did not and the word in question really had the shape *ov*, most likely it would have to be derived from an earlier *öv*. Thus, for our purposes, *öv* and *ov* can be reduced to the single form *öv*.

3.4. Revised table of historical forms

After applying the corrections and identifications discussed above, the revised table is as follows:

äv ~ äb: Orkhon
äv: Oghuz, Kipchak
äw: Karakhanid
iv: Kipchak
öj: Kipchak
öv: Armeno-Kipchak, Kipchak
üj: The Legend of Oghuz Khan dialect
üv: Karakhanid, Armeno-Kipchak

4. Analysis and conclusions

4.1. The suspicion of multiple roots

Before starting the actual reconstruction, there is an issue to be addressed first. Two types of forms of the lexeme in question have been interpreted in the past as deriving from different roots than the others. We will now examine whether these propositions are justified. Two hypotheses should be discussed:

1. Ramstedt's (1949: 56) proposition that the forms *öγ* are Mongolian loanwords.
2. Räsänen's (1949: 128–129) suggestion that the forms *uγ, ū* etc. (that is, forms with a back vowel and a velar consonant) continue the PT **ūk* 'tent ribs'.

4.1.1. The forms *öγ*

Although the unusual shape of the forms of the type *öγ* has been noted much earlier (cf. Grønbech 1979: 64–65), Ramstedt was the first to propose that they were a borrowing from the Mongolian *öb* 'inherited land or possessions, domain'. This was later challenged by Tatarincev (Tatarincev IV 354–355), although his argumentation consisted only of general remarks that such an adaptation of this Mongolian item is improbable. Sanžeev, Orlovskaja and Ševernina (2016: 220) failed to provide an etymology for the Mongolian *öb*, but take into consideration the possibility of it being somehow related to the Turkic lexeme under discussion without referring explicitly to the South Siberian forms. They also do not comment on Ramstedt's proposition.

Let us now analyze the relationship between the Mongolian *öb* and the Turkic *öγ*, starting with the semantics. Lessing (1960: 627) lists the meaning of the Mongolian *öb* as 'share, allotted part or portion, inheritance, property (to which a person has a legal title)'. However, the meanings in the Turkic languages are markedly different. In Tuvan, the primary meaning is 'yurt' (Tenišev 1968: 330), with the other meanings being evidently derived from it, in Soyot the meanings are 'tent, yurt, house, room' (Rassadin 2006: 103), and in Tofalar the meanings are similarly 'tent, yurt, house', although there are two additional meanings, namely 'circle' and 'ring (around the moon)' (Rassadin 2005: 76). A comparison of the Mongolic and Turkic semantics in isolation is rather inconclusive, with the possibility of imagining a transition from 'inheritance, property' to 'house, yurt', but at the same time, such a transition is not trivial. However, on the other hand, the meanings associated with the form *öγ* do not differ appreciably from the meanings found in other Turkic languages. Connecting the semantics of *öγ* to the Mongolian *öb* is possible, albeit somewhat far-fetched, but connecting it to the meanings of the Tkc. *äv*, *äb*, *üj* is much more convincing. This means that the semantics is an argument against Ramstedt's proposition and in favour of connecting *öγ* to all the other forms discussed in this paper.

As for the phonetics, Chabgataeva (2009: 55) fails to present any examples of similar adaptations of Mongolian loanwords in Tuvan. In her opinion, the Mongolic *-b* was adapted exclusively as *-p*, as in Tuv. *töp* 'centre' ← Mo. *töb* 'centre, middle, central etc.'. This is again an argument against Ramstedt's hypothesis. However, in terms of phonetics, the native origin of *öγ* is also doubtful. On the one hand, we have the correspondence Tuv., Tof. *öγ* : Tksh. *äv*, but on the other hand, Tksh. *dävä*, corresponds to Tuv. *tävä* (Tenišev 1968: 409) and Tof. *täbä* id. (Rassadin 2005: 101) and Tksh. *gäv-* 'to chew' corresponds to Tuv. *käγžän-* (Tenišev 1968: 234) and Tof. *keγžä-* id. (Rassadin 2005: 37).

These examples show the following correspondence:

Group **äv*		Group **öγ*
äv	:	*öγ*
äv	:	*äv ~ äb*
äv	:	*äγ ~ eγ*

However, we also have examples such as: Tksh. *öğür* 'group' and Tuv. *ȫr* (< *öγür*) 'id' (Tenišev 1968: 337), Tksh. *öğrät-* 'to teach' and Tuv. *ȫrät-* (< *öγürät-*) id. (Tenišev 1968: 337).

Here, the correspondences are

<div align="center">

Group *äv*   Group *öγ*

| *äv* | : | *öγ* |
| *öğ* | : | *öγ* |

</div>

As can be seen, the picture is somewhat complicated and it is difficult to compare the forms *öγ* with the rest. It seems that in terms of phonetics, both explanations of the origin of *öγ*, that is as native or as a Mongolian borrowing, are problematic and rather unconvincing. The semantics point towards a native origin, albeit not unequivocally. I believe that at the present moment, we have to accept that this issue is still unsettled and requires further study.

4.1.2. The forms *uγ, ū*

The most unusual aspect of the forms *uγ, ū* is their back vowel. Forms with a back vowel can be found in several languages, but can mostly be explained as late, secondary developments, as in, e.g., Uzbek. The forms *uγ, ū*, with which we are dealing here are problematic, as it is difficult to propose any regular sound laws that could account for them. On the other hand, it should be remembered that they come from poorly attested Siberian dialects, mostly from Radloff's dictionary. The identification of the dialects in Radloff's work is not always entirely clear (cf. e.g. K. Stachowski 2021), and these dialects lack comprehensive grammatical descriptions, so the existence of certain processes resulting in backing of the vowels cannot be ruled out.

However, a different explanation was also put forward by Räsänen (1949: 128–129). He believed that these forms do not continue the same root as e.g. the Tksh. *äv*, but come from the PT **ūk* 'tent ribs' (ÈSTJa I 583). Phonetically, this seems to be a plausible hypothesis. Semantics is also not problematic, as similar metonymies, i.e. 'part of the house' > 'house as a whole' are common in many languages across the world, cf. e.g., Trkm. *tām* 'house' < PT **tām* 'roof, wall' (Łukasik 2022: 20). For these reasons, I am inclined to agree with Räsänen and treat these forms as being unrelated to the Tksh. *äv* and its cognates.

4.2. Formal reconstruction

Having examined the previous etymologies and analyzed both the modern and the old forms, we can proceed to the etymology. Almost all the authors working on the etymology of the lexeme in question have proposed proto-forms of the type *EB*, that is an illabial front vowel and a labial consonant (cf. Table 1). Thus, it appears that the main problem we have to resolve is which illabial front vowel and which labial consonant should be reconstructed. However, I believe that we should not *a priori* exclude the possibility of a different reconstruction, and because of this, we shall start with a systematic analysis of the forms with labial vowels.

First, we will compare the groups *äv and *üj. Tksh. äv corresponds to Bash. öj (< üj) and Kaz. üj, as in Tksh. säv- 'to love', Bash. höj- (Uraksin 1996: 738), Kaz. süj- id. (Shnitnikov 1966: 184), as well as Tksh. gäv- 'to chew', Bash. köjš- id. (Uraksin 1996: 296), Kaz. küjis 'chew' (Shnitnikov 1966: 130). However, we also find Bash. öjköm 'cluster, pile' (Uraksin 1996: 480) (< öj-), and Kaz. üj- 'to amass, to put in a heap (Shnitnikov 1966: 285), both corresponding to Tksh. öjük ~ höjük 'burrow, grave' < *öj- (KEWT 274; ÈSTJa I 620–621). Similarly, Bash. böjök 'big', höjle- 'to speak' (Uraksin 1996: 102, 737) correspond to Tksh. büjük and söjle-.

The adduced examples illustrate the following correspondences:

Group *äv		Group *üj
äv	:	üj
üj	:	üj
öj	:	üj

It is impossible to establish the context in which *üj might have split into three different continuants in the group *äv. Because of this, it is much more likely that the group *üj has merged three sound groups which initially were distinct. This in turn means that *üj could not have been the Proto-Turkic form of the lexeme in question.

When comparing the groups *äv and *üv, we find the correspondence äv : üv ~ iv in examples such as Tksh. čävrä 'circle', SWK civrä, NWK č'uvr'a id. (Baskakov et al. 1974: 613, 642), and Tksh. säv- 'to love', SWK siv-, NWK s'uv'- id. (Baskakov et al. 1974: 472, 500). However, we also find examples such as Tksh. čivi 'nail' and SWK civ 'capital (of a column)', NWK č'uv 'hook' (Baskakov et al. 1974: 613, 642), as well as Tksh. güve 'moth' and SWK givä id., NWK g'uv'a id. (Baskakov et al. 1974: 160, 167).

The adduced examples illustrate the following correspondences:

Group *äv		Group *üv
äv	:	üv
iv	:	üv
üv	:	üv

As the situation is analogous to that in the previous paragraph, the same conclusion has to be proposed, that is that *üv could not have been the Proto Turkic form of the lexeme in question.

The sound group äv of the group *äv has two correspondences in Uyghur. Antevocalically äv equals Uyg. ög, as in Tksh. dävä 'camel', Uyg. tögä id. (Necip 1995: 423). Before consonants and in the auslaut äv corresponds to Uyg. öj., as in Tksh. säv- 'to love', Uyg. söj- 'to kiss' (Necip 1995: 363), Tksh. gäv- 'to chew', Uyg. köjši- id. (Necip 1995: 209).

However, Uyg. öj and ög also correspond to öj and öǧ, as in; Uyg. sögät 'willow' (Necip 1995: 361), as well as Tksh. söǧüt id., Uyg. öget- 'teach' (Necip 1995: 304), Tksh. öǧret- id., Uyg. köjnäk 'shirt' (Necip 1995: 209), Az. köjnäk id. (Hacıyev 2005: 445).

The adduced examples illustrate the following correspondences:

Group *äv		Group *öj
äv	:	*öj, ög*
öj	:	*öj*
öǧ	:	*ög*

The situation is parallel to the previous paragraphs, which again leads us to conclude that **öj* could not have been the Proto-Turkic form of the lexeme in question.

The case of Turkmen is problematic, as the phonetic correspondences between Turkmen and other languages are much more complicated than those presented above. If we compare Turkmen to the group **äv*, we find examples such as: Tksh. *säv-* 'to love', Trkm. *θöj-* id. (Baskakov et al. 1968: 589), Tksh. *dävä* 'camel', Trkm. *Düje* id. (Baskakov et al. 1968: 289), Tksh. *čävrä* 'circle', Trkm. *čöwre* 'inside out, the inside, wrong side of clothes' (Baskakov et al. 1968: 737), Tksh. *gäv-* 'to chew', Trkm. *gāwüš* 'chew' (Baskakov et al. 1968: 239). However, we also find examples such as Tksh. *öǧlä* 'noon' (in which the *ǧ* might not be etymological, cf. KEWT 271–272), Trkm. *öjle* 'afternoon', Az. *köjnäk* 'shirt' (Hacıyev 2005: 445), Trkm. *köjnek* id. (Baskakov et al. 1968: 411).

These examples illustrate the following correspondences:

Group *äv		Turkmen
öj	:	*öj*
äv	:	*öj*
äv	:	*ā̆w*
äv	:	*öw*
äv	:	*üj*

It is difficult to reconcile the Turkmen data with the data from other languages, a fact that can be interpreted in different ways. Perhaps these correspondences point towards four separate Proto-Turkic sound groups which have been merged into *äv* in the group **äv*. However, the previous comparisons show that these four sound groups could not have simply been **öj*, **äv*, **öv*, and **üj*. This means that if we were to accept Turkmen as preserving the PT state, we would have to propose at least a few completely new sound groups (and, therefore, new phonemes). It goes without saying that such *ad hoc* creations of new entities should be avoided unless completely necessary. The other option is to assume some inner-Turkmen processes which could explain the splitting of a single PT sound group, although at the present moment no such explanation presents itself. Finally, explanations that fall between these two extremes should also be taken into consideration. Perhaps some of these four Turkmen forms result from inner-Turkmen splitting, while others require a new PT phoneme in their etymon. This is without doubt a complicated issue which requires further, more detailed study. At present, the issue must be put aside. However, it must be stressed that its resolution in the future might impact to a considerable extent upon the conclusions arrived at in this paper, perhaps even requiring their complete revision.

As we have seen, with the important exception of Turkmen and the group *öγ, all the types of forms with a labial vowel appear to be secondary.[26] What remains are the groups *äv and *äb. This agrees with the historical data, as the oldest attestations are all of this type. Before we discuss the consonant, let us first consider the vowel. The above considerations show that in all likelihood this must have been an *E* type vowel. Some researchers have attempted to reconstruct a narrow *e* (cf. Table 1), but unfortunately their reasoning was not made clear in their works. Let us take a look at the available data and see whether such a reconstruction is justified.

The historical sources do not provide any forms with a narrow *e*. The only non-labial form that could suggest a different reconstruction than *ä is the Kuman form *iv*. However, it can also be explained without reconstructing a narrow *e*, as it has been postulated that the Kuman language also took part in the Volga Vowel Shift (Tekin 1987). While this process is usually associated with the labial vowels, the raising of *ä to *i* is also a characteristic feature of the Volga Turkic Languages (Dmitriev 1948: 12–14). In this light, Kuman *iv* could simply be derived from an earlier *äv.

Another factor that might indicate a narrow *e* is the Kipchak form *üj*. Those researchers who have reconstructed a broad *ä*, had attempted to derive *üj* from *öj*, thus implying that the development consisted of two phases. The first was the labialization and the second was the raising of the low vowel. This allows the necessity of reconstructing a vowel of intermediate height to be avoided. Such a development is difficult to justify in contemporary languages, as the only examples of *öj* in Kipchak, that is in Tatar and Bashkir, are clearly secondary. However, the development *öj* > *üj* can be postulated for Proto-Kipchak.[27] The raising of *ö* to *ü* in the position before *j* can be regularly observed in all the Kipchak languages, and occurs in words going back to different PT forms. The examples include: Tksh. *öğrät-* 'to teach', Bash. *öjrät-* id. (Uraksin 1996: 482), Kaz. *üjrät-* id. (Shnitnikov 1966: 286); Tksh. *öğür* 'group', Bash. *öjör* 'herd' (Uraksin 1996: 481), Kaz. *üjir* 'herd of horses, flock crowd, society' (Shnitnikov 1966: 286). Thus, the forms *üj* really should be derived from *öj, and they cannot be used as an argument in favour of a narrow *e*.

The third argument is the Azeri form *ev*. The usual argument for reconstructing the PT *e consists of comparing Azeri and Chuvash (M. Stachowski 1998: 162–164). The equation is as follows:

PT *e > Chuv. *i*, Az. *e*
PT *ä > Chuv. *a*, Az. *ä*

Although Chuvash does not have a cognate of *äv*, it does have a derivative *avlan-* 'get married' (Skvorcov 1982: 22). Thus, the Chuvash and Azeri data do not coincide.

---

[26] Regrettably, the group *üγ cannot be analyzed here, as the poor attestation of the languages which are part of this group does not allow for satisfactory comparisons. Nevertheless, the chance that this group might alter our conclusions is not great.

[27] It is not important within this context whether we consider Proto-Kipchak an actual uniform language, or an umbrella term for a certain period in the development of the Kipchak dialects.

Therefore, overall, the arguments in favour of reconstructing a narrow *e* in the proto-form of the word in question are not very convincing. The available data strongly point towards **ä* and this reconstruction will be accepted here.

Let us now proceed to the consonant. The forms we have to compare are the pre-forms **äv* and **äb* plus the Old Turkic forms *äw, äv* and *äb ~ äv*.

The **äb* group is in fact composed of two sub-groups. The majority of the languages in the **äb* group have merged the PT **b* with **p*, as they have developed very strict phonotactic rules governing voicedness, only seldom allowing phonological opposition between voiced and voiceless consonants (cf., e.g., the situation in Khakas: Baskakov 1975: 29–30). For these languages, the pre-form could have been either **äb*, or **äp*. The second sub-group only includes Shor, which has *äm*, whose pre-form could have been either **äb* or **äm*. The common source of these two sub-groups must have been **äb*, which also points towards Proto-Turkic **äb*.

The pre-form **äv* can be equated with the historical Oghuz form *äv*. In the Oghuz languages, in native words the consonant *v* comes from the PT **b* in a post-vocalic position (Doerfer 1975–1976a: 102, 126). It has become a separate phoneme as a result of several processes, namely, the emergence of a secondary *b* from the intervocalic *p* after a long vowel, the infusion of a large number of loan words with *v* not limited to the post-vocalic position, and the emergence of an initial *v* in the so-called "four words" (Doerfer 1976). Thus, the form *äv* unequivocally points to the reconstruction **äb*.

The consonant in the Karakhanid form *äw* differs phonetically from the Oghuz *v*, but appears in the same positions (MK I 55), so there is no need to discuss it separately.

The form *äv ~ äb*, from the runic script, has been discussed above, but let us repeat those remarks, as they are vital to the reconstruction. The transcription *äb* is the traditional rendering, following the notation of V. Thomsen. Other scholars have later suggested that post-vocalically this sign should be read as *v* (Erdal 2004: 63–67). This proposition was based on the evidence from the Old Turkic texts written in scripts other than the runic. Whatever the correct reading, however, it has no impact on the phonological system of Orkhon Turkic, as this *v* would still be only a combinatory variant of /b/.

The conclusion emerging from this analysis is that the Proto Turkic consonant has to be phonologically reconstructed as */b/*. Judging by the situation in Old Turkic, it is possible and perhaps even likely that allophonic spirantization of the post-vocalic *b* into *v* or *β* might have already occurred at the Proto-Turkic stage. Regrettably, we lack the tools necessary to establish the exact pronunciation with certainty.

In summary, the Proto Turkic form of the lexeme in question has to be reconstructed as **äb*.

4.3. Development from PT to the pre-forms and problems connected to it

Having reconstructed the Proto Turkic form, the ways in which the lexeme has developed in different groups and languages should be considered. As the development from the pre-forms to the modern languages has already been examined in

section 2, in this section we will attempt to sketch a model linking the Proto Turkic form to the pre-forms:

PT *äb > 1. äv, 2. *äb, 3. *öj > *üj, 4. *üv, 5. *öj, 6. *öɣ, 7. *üɣ

The two Siberian Tatar pre-forms were not included in the above model for the reasons stated in 2.17. The group äv resulted from the spirantization of *b after vowels and the subsequent phonologization of v. The group äb seems to continue the Proto-Turkic stage unchanged, but this is somewhat superficial, as the phonological status of b in this group is different from Proto-Turkic. The Kipchak *üj forms can be derived from an earlier *öj, which could perhaps be derived directly from *äb. However, a question arises in this context, namely, what is the relation of this group to the group üv? Is the Western Karaim form *üv more archaic than the Kipchak üj, or is it secondary? During the discussion of the narrow e, we have concluded that the raising of the vowel in *üj was caused by the adjacent j. This would suggest that the form *üv must be younger than *üj, but on the other hand, it is easier to propose *b > *v > *j, than *b > (*v) > *j > *v. This issue requires further study. The next group raises similar questions. What is the relation of the pre-form *öj to the historical Karakhanide form üv? It has also yet to be explained how the forms *üɣ and *öɣ arose. The curious relation between b and g in the Turkic languages is a long-standing problem in Turkology, and this issue is another in which it manifests itself (Grønbech 1979: 64–75). All these numerous questions still await a solution.

4.4. Semantic problems

The semantics of the word in question were mostly left out of the present discussion because in fact they appear to be quite straightforward. The main meanings which *äb assumes are all connected to a dwelling place. Depending on the language, it can be the general term for 'house', or a more specific word for a type of house, most commonly a yurt. These meanings can be in all likelihood also attributed to Proto-Turkic. The question of whether the original meaning was the general notion 'house', or the specific notion 'yurt', which has been raised occasionally, appears futile to me. Such a discussion would only make sense if we could point to an opposition with a different lexical item, in which one part of the opposition assumed one meaning, and the other the second meaning. However, it appears that such an opposition cannot be reconstructed for PT.

A more important problem is how Tatarincev's proposition of deriving the verb *äbir 'turn, twirl' from *äb 'house' (cf. section 1.11.) should be evaluated. Such a connection would, in Tatarincev's eyes, require a modification of the semantic interpretation of *äb. As the meaning 'to turn' is hardly derivable from 'house', the original meaning of this word must had been 'curved, round'. The semantic evolution would be as follows: 1. 'round' > 2. 'a round building' > 3. 'yurt' > 4 'a house in general'. To strengthen his argument, he adduces the Tofalar form of the word in question, öɣ, which, apart from the meaning 'house', also has the meanings 'circle, ring, wreath

around the moon'. This is another issue which can only be brought to attention in this article, but which requires a separate study. The forms of *äbir* in different languages generally agree phonetically with the forms of *äb*, but this is not enough to come to any conclusions. What is needed is a study of the secondary meanings of *äb*, to see if there are meanings which, on the one hand, cannot be derived from 'house', and on the other hand, can be connected to the notion of 'being round' or to 'turning' and 'twirling'.

4.5. Summary

This article does not claim to be the definitive study of *äb*. As we have seen, some unresolved issues still remain, of which the most important is that created by the Turkmen data (cf. section 4.2). This data seem to contradict the data from other languages, which in turn raises questions about the validity of the proposed reconstruction. This issue requires further study. Nevertheless, it can be stated that the remainder of the data strongly points towards reconstructing PT *äb*. The reconstruction of a narrow *e* or a different labial consonant does not seem to be justified. Phonological comparisons have also excluded the possibility of reconstructing *üj*, *öj*, or *üv*.

Aside from the reconstruction itself, I hope that the possibly comprehensive discussion of the state of the art, and the collating and systematic treatment of the available data may prove helpful to researchers who might want to address this question in the future. Likewise, I also believe that a clear formulation of the questions still unanswered, as well as drawing attention to certain less well-known proposals, as in sections 4.1 and 4.4, is also important and constitutes a necessary step in the direction of conclusively solving the question of the origin and development of *äb*.

Abbreviations

Az. = Azeri; **Bash**. = Bashkir; **Chin**. = Chinese; **Chuv**. = Chuvash; **EK** = Eastern Karaim; **Ger**. = German **Jap**. = Japanese; **Kaz**. = Kazakh **Khak**. = Khakas; **Kum**. = Kuman; **Küä** = Küärik; **Leb**. = Lebedin; **LC** = Lower Chulym; **MC** = Middle Chulym; **MKor**. = Middle Korean; **NWK** = North Western Karaim; **OA** = Old Anatolian; **OC**. = Old Chinese; **OT**. = Old Turkic; **Ott**. = Ottoman; **PT** = Proto Turkic; **Sum**. = Sumerian; **SWK** = South Western Karaim; **Tat**. = Tatar; **Tksh**. = Turkish; **Tof**. = Tofalar; **Trkm**. = Turkmen; **Tuv**. = Tuvan; **Uyg**. = Uyghur; **Uzb**. = Uzbek; **WY** = Western Yugur

References

Ayazlı Ö. 2019. Sarı Uygurca Üzerine Etimolojik Denemeler: U–Ü Maddeleri. – *Türk Dili ve Edebiyatı Dergisi* 59.1: 1–26.

Awde N., Dirks W., Hikmatullaeva U. 2002. *Uzbek dictionary & phrasebook*. New York: Hippocrene Books.

Bang W., Rahmeti G.R. 1936. *Oğuz Kağan Destanı*. İstanbul: Burhaneddin Basimevi.

Baskakov N.A. 1956. *Russko-nogajskij slovar′*. Moskva: Gosudarstvennoe izdatel'stvo ino-strannych i nacijonal'nych slovarej.

Baskakov N.A. 1964. *Russko-altajskij slovar′*. Moskva: Izdatel'stvo "Sovetskaja ènciklopedija".

Baskakov N.A. 1967. *Russko-karakalpakskij slovar′*. Moskva: Izdatel'stvo "Sovetskaja ènciklopedija".

Baskakov N.A. (ed.). 1975. *Grammatika chakasskogo jazyka*. Moskva: Izdatel'stvo "Nauka".

Baskakov N.A. 1985. *Dialekt lebedinskich tatar-čalkancev (kuu kiži). Gramatičeskij očerk, teksty, perevody, slovar′*. Moskva: Izdatel'stvo "Nauka".

Baskakov N.A., Karryjev B.A. Chamzaev M.Ja. (eds.). 1968. *Turkmensko-russkij slovar′*. Moskva: Izdatel'stvo "Sovetskaja ènciklopedija".

Baskakov N.A., Zajončkovskij A., Šapšal S.M. (eds.). 1974. *Karaimsko-russko-pol'skij slovar′*. Moskva: Izdatel'stvo "Russkij Jazyk".

Baysähedow D. 2009. *Iňlisçe-türkmençe sözlük*. Aşgabat: Türkmen döwlet neşiryat gullugy.

Bektaev K. 2007. *Bol'šoj kazachsko-russkij, russko-kazachskij slovar′*. Almaty: Dolphin Agency.

Čankov D.I. 1961. *Russko-chakasskij slovar′*. Moskva: Gosudarstvennoe izdatel'stvo inostran-nych i nacional'nych slovarej.

Center for Turkic and Iranian lexicography and dialectology. [available online: https://ctild.indiana.edu/About/Home, accessed: 4 September 2024].

Chabgataeva B. 2009. *Mongolian elements in Tuvan*. Wiesbaden: Harrassowitz.

Chabičev M.A. (ed.). 1990. *Russko-karačaevo-balkarskij slovar′-razgovornik*. Čerkessk: Karačaevo-čerkeskoe otdelenie stavropol'skogo knižnogo izdatel'stva.

Clark L. 1998. *Turkmen reference grammar*. Wiesbaden: Harrassowitz.

Clauson G. 1957. Turkish *y* and related sounds. – *Studia Altaica: Festshrift für Nikolaus Poppe*. Wiesbaden: Harrasowitz: 33–45.

Clauson G. 1962. *Turkish and Mongolian studies*. London: The Royal Asiatic Society of Great Britain and Ireland.

Clauson G. 1972. *An etymological dictionary of pre-thirteenth-century Turkish*. Oxford: Clarendon Press.

CTILD = *Center for Turkic and Iranian lexicography and dialectology*.

Dmitriev N.K. 1948. *Grammatika baškirskogo jazyka*. Moskva, Leningrad: Izdatel'stvo Akademii Nauk SSSR.

Dmitrieva L.V. 1981. *Jazyk barabinskich tatar (materialy i issledovanija)*. Leningrad: "Nauka" Leningradskoe otdelenie.

Doerfer G. 1975–1976a. Das Vorosmanische (die Entwicklung der Oghusischen Sprachen von den Orchoninschriften bis zu Sultan Veled). – *Türk Dili Araştırmaları Yıllığı Belleten*: 81–132.

Doerfer G. 1975–1976b. Proto Turkic: Reconstruction problems. – *Türk Dili Araştırmaları Yıllığı Belleten*: 1–59.

Doerfer G. 1976. Die "vier Wörter" mit b- > v-, Null. – *Hungaro-Turcica*: 135–147.

DTS = Nadeljaev, Nasilov, Tenišev, Ščerbak (1969).

Dul'zon A.P. 1973. Dialekty i govory tjurkov Čulyma. – *Sovetskaja Tjurkologija* 2: 16–29.

Dyrenkova N.P. 1940. *Grammatika ojrotskogo jazyka*. Moskva, Leningrad: Izdatel'stvo Akademii Nauk SSSR.

Dyrenkova N.P. 1941. *Grammatika šorskogo jazyka*. Moskva, Leningrad: Izdatel'stvo Akademii Nauk SSSR.

EDT = Clauson (1972).

Erdal M. 2004. *A grammar of Old Turkic*. Leiden: Brill.

Eren H. 1999. *Türk Dilinin Etimolojik Sözlüğü. 2 baskı*. Ankara: Bizim Büro Basım Evi.

ÈSTJa I = Sevortjan (1974).

Ganiev F.A. (ed.). 1997. *Russko-tatarskij slovar'*. Moskva: "Insan".

Garkaveć O. 2000. *Urumskij slovnik*. Alma-Ata: Baur.

Garkavec A. 2010. *Kypčakskoe pis'mennoe nasledie. Tom III. Kypčakskij slovar' po armjano-pis'mennym pamjatnikam XVI–XVII vekov*. Almaty: Baur-Kasean.

Grønbech K. 1942. *Komanischer Wörterbuch. Türkischer Wortindex zu Codex Cumanicus*. Copenhagen: Einar Munksgaard.

Grønbech V. 1979. *Preliminary studies in Turkic historical phonology*. [transl. by J.R. Krueger]. Bloomington (IN): Indiana University Research Institute for Inner Asian Studies.

Gucu V., Vasiloglu K. 1996. *Slovar' v kartinkach*. Kišinev: "Lumina".

Guljamov Ja.G. 1957. Iz nabljudenij nad morfologiej taškentskogo govora. – Rešetov V.V. (ed.). *Materialy po uzbekskoj dialektologii*. Izdatel'stvo Akademii Nauk UzSSR: Taškent: 165–229.

Hacıyev E.İ. (ed.). 2005. *Azərbaycanca-İngiliscə lüğət. A–Z*. Bakı: "Şərq-Qərb".

Hamilton J. 1974. *Opla-/yopla-*, *uf-/yuf-* et autres formes semblables en turc ancien. – *Acta Orientalia Academiae Scientiarum Hungaricae* 28.1: 111–117.

Hamp E.P. 1976. Palatalization and harmony in Gagauz and Karaite. – Heissig W., Krueger J.R., Oinas F.J., Schutz E. (eds.). *Tractata altaica Denis Sinor sexagenario optime de rebus altaicis merito dedicata*. Wiesbaden: Harrassowitz: 211–213.

Hartmann M. 1904. Ein türkischer Text aus Kašgar II. – *Keleti Szemle* 5: 161–184.

Ischakov F.G., Pal'mbach A.A. 1961. *Grammatika tuvinskogo jazyka. Fonetika i morfologija*. Moskva: Izdatel'stvo vostočnoj literatury.

Judachin K.K. (ed.). 2000. *Russko-kyrgyzskij slovar'*. Biškek: Izdatel'stvo "Šam".

KEWT = Stachowski M. (2019).

Kononov A.N. 1958. *Rodoslovnaja Turkmen. Sočinenie Abu-L-Gazi chana chivinskogo*. Moskva, Leningrad: Izdatel'stvo Akademii Nauk SSSR.

Kononov A.N. 1960. *Grammatika sovremennogo uzbekskogo literaturnogo jazyka*. Moskva, Leningrad: Izdatel'stvo Akademii Nauk SSSR.

KS = *Qumuq til*.

Kurpeško-Tannagaševa N.N., Apońkin F.Ja. 1993. *Šorsko-russkij i russko-šorskij slovar'*. Kemerovo: Kemerovskoe knižnoe izdatel'stvo.

Lemskaja V.M., Kondijakov A.F. 2021. *Čulymskij jazyk. D. Pasečnoe tjuchtetskogo rajona krasnojarskiego kraja. 2007–2021 gg.* [vol. 1]. Tomsk: Izdatel'stvo tomskogo universiteta.

Lessing F. (ed.). 1960. *Mongolian-English dictionary*. Berkeley, Los Angeles: University of California Press.

Łukasik J. 2022. *Wyrazy w znaczeniu 'dom' w językach turkijskich. Próba analizy porównawczej i etymologicznej*. [unpublished BA thesis, Jagiellonian University].

Maḥmūd al-Kāšġarī. 1982. *Compendium of the Turkic dialects. Part I*. [ed. and transl. by R. Dankoff and J. Kelly]. Cambridge (MA): Harvard University Printing Office.

Malov S.E. 1957. *Jazyk želtych ujgurov. Slovar' i grammatika*. Alma-Ata: Izdatel'stvo Akademii Nauk Kazachskoj SSR.

Mardkowicz A. 1933. *Słownictwo karaimskie. Karaimsko-polsko-niemiecki słownik*. Łuck. [*sine nomine editoris*].

Menges K.H. 1982. Etymologika. – *Central Asiatic Journal* 26.1/2: 105–118.

MK I = Maḥmūd al-Kāšġarī (1982).

Musaev K.M. 1964. *Grammatika karaimskogo jazyka*. Moskva: Izdatel'stvo "Nauka".

Nadeljaev V.M., Nasilov D.M., Tenišev È.R., Ščerbak A.M. (eds.). 1969. *Drevnetjurkskij slovar'*. Leningrad: Izdatel'stvo "Nauka".

Necip E.N. 1995. *Yeni Uygur Türkçesi Sözlüğü*. [transl. by İ. Kurban]. Ankara: Bizim Büro Basımevi.

Németh M. 2014. A historical phonology of Western Karaim. The evolution of consonant harmony in the North-western dialect. – *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 131: 353–369.

Ölmez M. et al. 2008. *A study of the Middle Chulym dialect of the Chulym language*. Seoul: Seoul National University Press.

Polivanov E.D. 1929. Obrazcy ne-iranizovannych (singarmonističeskich) govorov uzbekskogo jazyka. – *Izvestija Akademii Nauk SSSR. Otdelenie Gumanitarnych Nauk*: 511–537.

Polivanov E.D. 1933. *Uzbekskaja dialektologija i uzbekskij literaturnij jazyk (k sovremennoj stadii uzbekskogo jazykovogo stroitelʹstva)*. Taškent: Uzbekskoje gosudarstvennoe izdatelʹstvo.

Polivanov E.D. 1934. K ètimologii tureckogo uj, öj 'dom', 'jurta'. – *Sbornik naučnych trudov UzNIIKS* I.2 [*Trudy lingvističeskogo sektora*]: 85–88.

Polivanov E.D. 1968. *Izbrannye raboty. Statʹi po obščemu jazykoznaniju*. Moskva: Izdatelʹstvo "Nauka".

Pomorska M. 2001. The Chulyms and their language. An attempt at a description of Chulym phonetics and nominal morphology. – *Türk Dilleri Araştırmaları* 11: 75–123.

Pritsak O. 1959. Das Karaimische. – Deny J., Gronbech K., Scheel H., Togan Z.V. (eds.). *Philologiae Turcicae Fundamenta*. Wiesbaden: Harrassowitz: 318–340.

Qumuq til. [available online: https://til.im/ru/slovar/russko-kumykskii-slovar, accessed: 4 September 2024].

R I = Radloff (1893).

Radloff W. 1882. *Phonetik der nördlichen Türksprächen*. Leipzig: T. O. Weigel's Verlag.

Radloff W. 1893. *Versuch eines Wörterbuches der Türk-Dialecte. Erster Band*. Sanktpeterburgъ: Tipografija Imperatorskoj Akademii Naukъ.

Ramstedt G.J. 1949. *Studies in Korean etymology*. Helsinki: Suomalais-Ugrilainen Seura.

Rassadin V.I. 2005. *Slovarʹ tofalarsko-russkij i russko-tofalarskij*. Sankt-Peterburg: Izdatelʹstvo „Drofa".

Rassadin V.I. 2006. *Slovarʹ sojotsko-russkij*. Sankt-Peterburg: Izdatelʹstvo "Drofa".

Räsänen M. 1949. *Materialen zur Lautgeschichte der türkischen Sprachen*. Helsinki: Societas Orientalis Fennica.

Rešetov V.V. 1957. Izučenie uzbekskich narodnych govorov. – V.V. Rešetov (ed.). *Materialy po uzbekskoj dialektologii*. Izdatelʹstvo Akademii Nauk UzSSR: Taškent: 5–30.

Rjumina-Syrkaševa L.T., Kučigaševa N.A. 1995. *Teleutsko-russkij slovarʹ*. Kemerovo: Kemerovskoe knižnoe izdatelʹstvo.

Róna Tas A. 1991. *An introduction to turkology*. Szeged: Universitas szegediensis de Attila József Nominata.

Roos M.E. 2000. *The Western Yugur (Yellow Uygur) language. Grammar, texts, vocabulary*. [unpublished PhD dissertation, Rijksuniversiteit te Leiden].

Sagidullin M.A. 2008. *Fonetika i grafika sovremennogo sibirskotatarskogo jazyka*. Tjumeń: Izdatelʹstvo "ISKER".

Sagidullin M.A. 2010. *Russko-sibirskotatarskij slovarʹ*. Tjumeń: Mandr i Kᵃ.

Sanžeev G.D., Orlovskaja M.N., Ševernina Z.V. 2016. *Ètimologičeskij slovarʹ mongolskich jazykov. Tom II. G–P*. Moskva: Institut Vostokovedenija RAN.

Ščerbak A.M. 1970. *Sravnitelʹnaja fonetika tjurkskich jazykov*. Leningrad: Izdatelʹstvo "Nauka".

Sevortjan È.V. 1974. *Ètimologičeskij slovarʹ tjurkskich jazykov*. [vol. 1]. Moskva: Izdatelʹstvo "Nauka".

Shnitnikov B.N. 1966. *Kazakh-English dictionary*. London, The Hague, Paris: Mounton & Co.

Široliev M.A., Oružov A.H. (ed.). 1951. *Russko-azerbajdžanskij slovarʹ*. Baku: Izdatelʹstvo Akademii Nauk Azerbajdžanskoj SSR.

Širaliev M.Š., Sevortjan È.V. 1971. *Grammatika azebajdžanskogo jazyka (fonetika morfologia i sintaksis)*. Baku: Izdatel'stvo "ÈLM".

Sjoberg A.F. 1963. *Uzbek structural grammar*. Bloomington: Indiana University / The Hague: Mounton & Co.

Skopeteas S. et al. 2011. *Urum basic lexicon*. [unpublished manuscript, University of Bielefeld].

Skvorcov M.I. (ed.). 1982. *Čuvašsko-russkij slovar′*. Moskva: Izdatel'stvo "Russkij Jazyk".

Stachowski K. 2009. The discussion on consonant harmony in Northwestern Karaim. – *Türkbilig* 18: 158–193.

Stachowski K. 2021. South Siberian material in Radloff's dictionary. – *Bitig Türkoloji Araştırmaları Dergisi* 1: 37–50.

Stachowski M. 1998. A minimal probabilistic development model of Proto-Turkic *E*-type vowels. – *Folia Orientalia* 34: 159–174.

Stachowski M. 2019. *Kurzgefaßtes etymologisches Wörterbuch der türkischen Sprache*. Kraków: Księgarnia Akademicka.

Subrakova O.V. (ed.). 2006. *Chakassko-russkij slovar′*. Novosibirsk: Nauka.

Tatarincev IV = Tatarincev (2008).

Tatarincev B.I. (ed.). 2008. *Ètimologičeskij slovar′ tuvinskogo jazyka. Tom IV. M, N, O, Ö, P*. Novosibirsk: Nauka.

TDK = *Türk Dil Kurumu Sözlüğü*.

Türk Dil Kurumu Sözlüğü. [available online: https://sozluk.gov.tr/, accessed: 4 September 2024]

Tekin T. 1987. Labial vowel shifts in Tatar and Bashkir. – *Tatarica. Studia in honorem Ymär Daher anno MCMLXX sexagenario*. Helsinki: Abdulla Tukain kulttuuriseura: 293–306.

Tekin T. 1994. Türk Dillerinde Önseste y- Türemesi. – *Türk Dilleri Araştırmaları* 4: 51–66.

Tekin T. 1995. *Türk Dillerinde Birincil Uzun Ünlüler*. Ankara: Simurg Kitapçılık.

Tenišev E.R. 1966. O jazyke Kyrgyzov ujezda Fujuj (KNR). – *Voprosy Jazykoznanija* 1: 88–96.

Tenišev E.R. (ed.). 1968. *Tuvinsko-russkij slovar′*. Moskva: Izdatel'stvo "Sovetskaja ènciklopedija".

Tenišev E.R., Todaeva B.H. 1966. *Jazyk želtych ujgurov*. Moskva: Izdatel'stvo "Nauka".

Toparlı R., Vural H., Karaatlı R. 2007. *Kipçak Türkçesi Sözlüğü*. Ankara: Türk Dil Kurumu.

Tryjarski E. 1969. *Dictionnaire arméno-kiptchak d'après trois manuscrits des collections viennoises. Tome I. Fascicule 3. X–O*. Warszawa: Państwowe Wydawnictwo Naukowe.

Tryjarski E. 2017. *Zapisy sądu duchownego Ormian miasta Lwowa za lata 1564–1608 w języku ormiańsko-kipczackim*. Kraków: Księgarnia Akademicka.

Tumaševa D.G. 1992. *Slovar′ dialektov sibirskich tatar*. Kazan': Izdatel'stvo Kazanskogo Universiteta.

Uraksin Z.G. 1996. *Baškirsko-russkij slovar′*. Moskva: Izdatel'stvo "DIGORA", Russkij jazyk.

Uraksin Z.G. (ed.). 2005. *Russko-baškirskij slovar′ v 2 tomach*. Ufa: Naučnoe izdatel'stvo "Baškirskaja ènciklopedija".

Useinov S.M. 2007. *Russko-krymskotatarskij slovar′*. Simferopol': "Tezis".

Verhoeven E., Moisidi V., Yordanoglu E. 2011. *Urum basic grammatical structures*. [unpublished manuscript, University of Bremen].

Yusup A., Memet P. 2013. Experimental analysis of Uyghur vowel umlauting. – *Proceedings of the 2013 International Academic Workshop on Social Science*: 782–785.

Zatov O. 1906. *Polnyj russko-tatarskij slovar′ (krymsko-tatarskago narečija)*. Simferopol': Tipografija Br. Ja. i M. Nutisъ.