

MAREK STACHOWSKI
Jagiellonian University in Kraków
marek.stachowski@uj.edu.pl
ORCID: 0000-0002-0667-8862

TWO APPROACHES TO ETYMOLOGICAL RESEARCH IN TURKIC LINGUISTICS

1

Three years after my etymological dictionary of Turkish (KEWT) appeared in print a review by Marcel Erdal (2022) was published. At first, it seemed to me a fortunate circumstance because I expected a valuable discussion with creative impulses. That was, unfortunately, not the case. The review is aggressive both in spirit and style.¹ Its author distorts the quoted content and reduces the entire Turkic etymology to equating modern Turkish words with their Old Turkic counterparts, according to the motto: What is absent from Old Turkic does not exist at all. No areal or comparative reflection is presented, no asterisks are used. Being not very polemical in nature, I was at first shocked but not ready to write a response. On reflection, however, I came to a different conclusion: Erdal's review article is a good opportunity to present two approaches to etymological research coexisting nowadays in Turkic linguistics which is a phenomenon undiscussed in our circles and hardly known in other philologies.

What follows below consists of two parts. First, in section 2, some accusations are adduced and discussed in order to present Erdal's way of thinking. I am not going to examine every claim and every word of the reviewer – a few examples suffice. Then, in section 3, conclusions are drawn and the titular two approaches are shown.

¹ Phrases such as “Stachowski is unaware of all this” or “... unaware that...” are commonplace in this review.

2

- (1) “He [= Stachowski] should have seen in Clauson 1972 that the source of *güvey* ‘bridegroom, son in law’ is not **küdeg* but *küdagü*.” (E 76)²

Or rather: Erdal should have seen that there are two entries rather than one in KEWT: *güvey* and *güveyi*, both meaning ‘bridegroom’. The former goes back to **küdeg*, the latter to **küdegü*. The difference between Clauson’s *d* and my δ is apparent because it is Erdal who writes <ɖ> here unlike Clauson who puts <ɖ̣> that means an interdental δ for him. Since the *e* ~ *ä* difference is only transcriptional (my <e> stands for an open vowel *ä*), my reconstruction *küdegü* is phonetically identical to Clauson’s form *küdagü*. In footnote 7 Erdal says as follows: “The sound in Khâqânî Turkic written with the letter *dhâl* belongs to the phoneme /d/; I will from here on not be referring to the Old Turkic phoneme as δ , as Stachowski does, but as *d*.” – but then also Clauson’s <ɖ̣> should be corrected (not “as Stachowski does” but, rather, “as Clauson and Stachowski do”). Anyway, Clauson also quotes the variant *küzägü*, that is, one with -z- which certainly better matches - δ - than -*d*-.

Besides, Erdal has not only changed Clauson’s form but he also associated Clauson’s *küdagü* with my entry *güvey* instead with *güveyi*. In other words: he takes a wrong variant and equates it with a wrong word.

What is, however, much more important is the fact that Clauson does *not* reconstruct anything; he just adduces Old Turkic forms and, at best, shows their Old Turkic structure like saying that *küdagü* comes from *küda-* and, say, *kedil-* is a passive form of *ked-*. In KEWT 165 the word *güvey(i)* is ultimately derived from a Common Turkic **küδ* whose possible original meaning was ‘brother-in-law’.

- (2) “Stachowski thinks that ‘yesterday’ is the primary meaning of the lexeme *dün* (< *tün*) and ‘night’ its secondary meaning, with the argument that this latter is in use in marginal areas and not in the center of the Turkic world; however, it is typical for innovations to emerge in the center, leaving archaicities to the margin. The idea would have hardly come up if the author knew of *tünäk* ‘dark place, prison’, *tünär-* ‘to be or get dark’, the common *tünäriḡ* ‘dark’ or *tünlä* ‘at night’.” (E 77)

The idea that innovations “emerge in the center, leaving archaicities to the margin” is hopelessly obsolete. Dialectologists can recall dozens of examples of innovations emerging far away from the centre, contrary to Matteo Bartoli (1873–1946) who was the first linguist to formulate that rule – but he published his opinions in the late nineteenth and early twentieth century and one can hardly believe without surprise that a linguist seriously uses Bartoli’s claim in the 21st century. It is true, changes introduced in peripheries are mostly unknown in the centre; nevertheless, they do exist. Peripheries are no less innovative than the centre.

² The word *Turkish* in the title of Clauson (1972) stands for ‘Turkic’ because this author never accepted the derivative *Turkic*.

But the problem is that Erdal's report is unreliable. In the entry under discussion, I did not contrast the centre and the peripheries but, rather, the entire area and the peripheries. I wrote, the meaning 'yesterday' is known to *all* Turkic languages (lit. "in der ganzen Turcia") while the meaning 'night' only occurs in Khakas, Tuvian and Tofalar in the East as well as in Turkish and Crimean Tatar in the West, i.e. only in peripheries. It seems therefore comparably realistic (lit. "vergleichbar realistisch erscheint daher...") that the semantic evolution was going in the opposite direction: 'yesterday' > 'last night' > 'night'. I did not claim anything, I only allowed a possibility different from the traditional one. The areal distribution is an important indication that the semantic change might have started exactly in outlying areas.

Moreover, enumerating four derivatives with the meaning 'dark' does not prove the primary character of that meaning for the root. Even in the eyes of freshmen.

- (3) "Stachowski derives *büyü* 'magic' from hypothetical **büg-* 'to know', which is said to be the base of Old Uygur *bügün-* 'to know' and *bügüş ~ bögüş* 'knowledge'. However, we know *at least* since Wilkens 2021 that Old Uygur *bökün-* and *böküş* and their base *bök-* have /k/. (E 77)

First, KEWT was published in 2019 so that I can hardly be reproached for not having taken account of Wilkens (2021). Secondly, Erdal leaves the fact unsaid that also Turkish dialectal forms such as *büğ* and *bügü* are adduced in KEWT which makes the existence of a Common Turkic **g* realistic, even though it differs from Old Turkic. Turkish *-g(-)* cannot possibly be explained as a descendent form of the Old Turkic *-k(-)*.

- (4) "As the entries *doğmak* and *doğru* show, Stachowski is unaware that Common Turkic *tug-* 'to be born; (of the sun and moon) to rise' and *tog-* 'to cross, traverse' are two distinct verbs (see e. g. Clauson 1972: 465). *doğru* comes from the converb of the causative of Common Turkic *tog-*, not *tug-*." (E 77)

The essence of this passage is: Turkish *doğru* < Common Turkic *tog-* (I would put an asterisk before a reconstructed form). What I wrote in KEWT goes as follows: "**doğru** 'gerade(aus), direkt' [...] < **toguru*, Ger. [= Gerundium = converb] < **togur-* [...] (Kausat. < **tog-* [...])." Where did Erdal find a *tug-* here? Where, indeed? – It is only under *doğmak* 'to be born' that I made a following reference in parentheses: "([...] für die Lesung *tug-* s. TMEN III 201)."

- (5) "*akçe* 'money' is, on the other hand, likely to come from Common Turkic **āk* 'white', an etymology accepted by venerable scholars such as Tietze 2002. The author dismisses it because of its Ottoman spelling with *ghain*, apparently unaware of consonant voicing regular in Oguz languages after long vowels." (E 80)

Barring the argument from authority (I shall not discuss whether Tietze is one in etymology)³ Erdal once again blames me for being unaware³ of something. What he

³ Tietze also derives a Persian loanword in Turkish *çene* 'jaw, chin' from a fanciful Common or Proto-Turkic (he does not name the language) **iç-ejek* 'inner jaw' in his dictionary (does an

gives as a rule, however, is only the *first* half of the rule. Erdal is “apparently unaware” of its second part. It has been known since Vilhelm Grønbech (1902: 60, § 80), that is for 120 years, that Proto-Turkic unvoiced stops turn voiced in Turkish if *two* conditions are fulfilled: a vowel preceding the stop was originally long *and* the stop itself was followed by a vowel, that is $*-\bar{V}kV- > -VgV-$. Erdal’s example is the Turkish word *ağar-* ‘to turn white’; this is a very good example because the stop stands after an original long vowel and before another vowel, that is (*ağar-* <) *agar-* < $*\bar{a}kar-$. Which is not the case with *akçe* – no vowel follows the stop *k* in *akçe* and the stop cannot, thus, be voiced. That is why the old etymology should necessarily be rethought, even though Tietze adduces it. It was eighteen years ago that the problem was successfully solved by Hatice Şirin User (2004: 132–135) who derived the form *akçe* ‘money’ from another word, namely from $*agi+ça$ < Old Turkic *agi* ‘silk; treasure’. I have accepted Şirin User’s etymology which also explains older Turkish notations such as *agça* (< $*agiça$) and I fail to see why Erdal has not mentioned her in this context which she certainly deserves. A possible substantive discussion of the word would be much better than leaving unsaid the second half of the rule and reproaching me for ostensible unawareness of the first half.

- (6) “A verb *āg-* ‘to get toppled’ in use in Turkmen but not attested in Old Turkic is brought in to account for both *agir* ‘heavy, difficult’ and *aksak* ‘lame’. The semantic connection looks unlikely for both lexemes but, more importantly, there is no deverbal suffix *-sA-*; also, the vowel of the proposed source is long whereas the vowel of *aksa-* ‘to limp; to function improperly’ is short.” (E 80)

A few problems are involved in this passage. Firstly, “a verb *āg-*” (again, why without an asterisk?) is not brought in to account for *aksak* ‘lame’ for the simple reason that there is no entry *aksak* in KEWT.

Second, the fact that “[t]he semantic connection looks unlikely” is nothing but the reviewer’s individual feeling. If English *start* ‘to jump up, move one’s body suddenly’ etymologically equals the German verb *stürzen* ‘to tumble down; to overthrow’ and they both are accepted to be descendants of an Indo-European verb $*(s)tere-$ ‘to be stable, immobile’ of which one of the derivatives is Scottish English *strunt* ‘to walk proudly’ while another derivative is Middle High German *strunz* ‘stump, fragment, splinting’ (Levickij 2010: 513 sub STERT-) one is ready, indeed, to accept the meanings ‘to get toppled’ and ‘heavy’ (or ‘lame’) as related.

Third, the claim that there is no deverbal suffix *-sA-* is valid for Old Turkic but such a suffix occurs in Turkish dialects (see Güzel 2019: 613–614), a fact once again ignored by Erdal.

Fourth, Erdal says “the vowel of *aksa-* [...] is short”. In what language? In Turkish? All original long vowels were shortened in Turkish. The only Oguz language that has preserved original long vowels is Turkmen and the Turkmen form of the verb

‘outer jaw’ exist?) – should we accept it because Tietze wrote it even though it is phonetically impossible and semantically weird, to say the least?

is exactly *āg-* which clearly points to a Proto-Turkic long **ā-*. Too bad, Erdal passes over in silence that the Turkmen form has been adduced in KEWT as a reason for reconstructing a long vowel.

- (7) “The worst etymology I have come across is, I think, that of *evirmek* ‘to change, invert, alter, to turn over or around’, which is derived from **eb* ‘house’ through **ebür-*; the original meaning of the verb would then, according to Stachowski, be ‘to go around the house.’” (E 82).

Erdal’s report is misleading again. I was quite happy with the formal reconstruction of the Proto-Turkic form **ebür-*. Then, however, I also added, in the brackets, the following piece of information (here in English translation): “(ËSTJa I 498: ? < **eb* ‘house’; the original meaning should, thus, have been **‘to go around the house’*.” As we see, I reported on what had been suggested, with a question mark, in ËSTJa.

- (8) “The author has written a textbook on etymology, which has even been translated into Turkish (Stachowski 2011). It was therefore doubly important to get into all this detail here, to warn the uninitiated user to be careful: The review of such a work can’t possibly mention everything.” (E 82)

That is the closing passage of Erdal’s review. It is imprecise. The textbook was not translated into Turkish because “the author” wrote it himself in Turkish and then asked Professors Bülent Gül and Faruk Gökçe to check his Turkish (see the final passage in Stachowski 2011: viii) which is a common practice.

The general overtone of Erdal’s lines is that of a warning for the Turkish uninitiated: reading Stachowski’s books is a very risky job. Why should a Turkish reader mistrust Erdal’s words? Maybe because Turkish reviews of the textbook (Yıldız 2011; Hirik 2012; Öztürk 2012) were very positive? But today, ten years later only a few students will have access to those reviews. I would, thus, suggest the Turkish readers should first read *Etimoloji* and then decide on their own if it was a risky job.

In this context, a marginal but notwithstanding surprising fact should be mentioned that Erdal does not quote *Turkish* etymological works in his review devoted to *Turkish* etymology. Not a single one. And this despite his perfect command of Turkish. I can assure everyone that numerous valuable publications can be found in Turkish and ignoring them is a huge mistake.

Similarly, Erdal disregards publications in Russian. But how can one, then, work on Turkic languages if the majority of them are spoken in Central Asia, that is, on the territory of the former USSR, and the literature concerning them is mainly in Russian? In his review, Erdal only cites ËSTJa – nowadays the biggest collection of comparative word material and etymological ideas (seven volumes appeared up to now; the eighth is ready but not yet published). Erdal, however, calls ËSTJa “that compilation of utter ignorance.” (E 82)... I really do not know what to say.

I am not going to continue this discussion with M. Erdal.

3

Let us now turn to two approaches. As can be seen above, Erdal only respects two authors besides himself: Gerard Clauson and Andreas Tietze. They both were philologists rather than linguists⁴ and wrote etymological dictionaries without proper etymologies. These words are no malevolent criticism. Let us take an example, *un pour tous*:

Clauson (1972: 45b): “*adıg* ‘bear (animal)’. S.i.a.m.l.g. [= survives in all modern language groups] in various forms, usually *ayı/ayu* [...]” – That is all, apart from a handful of examples from the period from the 8th to the 14th century.

Tietze (2002: 244b; English added by me): “*ayı/ayı* ‘bear’ < Old Turkic *adıg* id.” – That is all, apart from one Turkish phrase from a 16th century text and two phrases from 20th century novels.

Are these etymologies of venerable scholars? These authors generally only combine words with their Old Turkic correspondences which is an aim of a comparative dictionary rather than that of an etymological one and they do not go beyond simple rapprochements. An etymology is expected to explain the original form and meaning of the given word.⁵ Neither of these is shown by Clauson and Tietze.

The same word in KEWT (in English):⁶

KEWT 69: “*ayı* biol. ‘bear (*Ursus*)’ = Turkmen *ajy* id. = Khakas *azyg* id. = Old Turkic *adug* id. < **adug* ‘old’ (taboo word for a bear) < **aδ-* ‘to grow older, be getting old’ > Teleut-Oirat *aj(y)k-* ‘to become aged, get much older’” [bibliography follows].⁷

One can discuss my etymology, it is true, but no one can discuss with Clauson or Tietze because they offer no etymology at all.

Another feature typical of the Clauson – Tietze – Erdal triad is their treating Old Turkic as a source of Turkish. And what is Old Turkic? Erdal (1998: 138) explains the notion as follows: “[...] Old Turkic is taken to be the language underlying three corpora. The first one consists of [...] inscriptions [7th-10th c.] [...] in the

⁴ Clauson’s way of reasoning can be understood from a passage in his book: “As a young man I had always accepted the theory that the Turkish and Mongolian languages were genetically related. [...] [S]o when a Romanized text of the *Secret History of the Mongols* [...] became available, I tried to read it. I did not begin to understand it, and I could find nothing Turkish about the language in which it was written. And so I came to the conclusion that [...] the Altaic theory [...] was almost certainly wrong [...] (Clauson 1962: XIII). – Roy Andrew Miller (1991: 224) called that fact “[...] Sir Gerard’s shocking discovery that merely because he knew [20th century] Republican Turkish he could not by that fact alone read [13th century] Middle Mongolian without first learning the language [...].”

⁵ That was already meant by Chrysippus of Soli when he coined the term *ετυμολογία*.

⁶ In non-Turkish Turkic words: KEWT ⟨j⟩ = Turkish ⟨y⟩; KEWT ⟨y⟩ = Turkish ⟨ı⟩ (as in English *Kyrgyzstan* = Turkish *Kırgızistan*).

⁷ Old Turkic covers different dialects in huge territories (see below). It is, thus, natural that various phonetic variants occur in Old Turkic texts. In DTS *adug*, *adyg*, *aδyg*, and *ajyg* are attested.

territory of the [...] present-day Mongolia [...] and the Yenisey basin. The second and most extensive corpus consists of [...] Old Uyghur manuscripts [9th-13th c.] from northwest China [...]. The third corpus consists of eleventh-century texts from the Karakhanid state [present-day Uzbekistan and Kyrgyzstan].” Three territories, three traditions, three scripts (runic, Uyghur, Arabic), three distinct compositions of tribes. Covering them all in one umbrella term results from a scholarly convention, additionally reinforced by the ambiguous Turkish term *Türkçe* ‘1. Turkic; 2. Turkish’ that makes it possible to understand *Eski Türkçe* ‘Old Turkic’ as ‘Old Turkish’. In reality, Old Turkic was no direct source of Turkish. The ratio of Old Turkic to modern Turkish is approximately that of Old Scandinavian to modern German. At one place of his review Erdal says: “Some features of Turkish are archaic, even more archaic than Old Uyghur.” (E 77) – can a modern language be more archaic than its source?

What has been said above is one approach to etymological research in Turkology. It does not seem to have new young entrants.

The other approach was initiated by Willi Bang, the creator of Turkic historical-comparative linguistics in the early 20th century. Its most typical feature is combining data from historical sources *and* modern languages and dialects. Bang made his etymologies this way even though he had at his disposal only few sources for Central Asian languages or Anatolian dialects. He never confined himself to Old Turkic only. His followers did not either. I am just a link in a long chain of Bang’s spiritual pupils, to name but a few (in chronological order of birth years): M. Räsänen (Finland), N. Poppe (Russia, USA), Gy. Németh (Hungary), K. H. Menges (Germany, USA, Austria), H. Eren (Turkey), G. Doerfer (Germany), A. M. Ščerbak (Russia), T. Tekin (Turkey), A. V. Dybo (Russia). Are they all no venerable scholars just because they allow for modern and dialectal data and try to explain the original form and meaning instead of being happy solely with rapprochements with Old Turkic?

References

- DTS = Nadeljaev V. M. 1969
 E = Erdal 2022
 ÈSTJa = Sevortjan 1974
 KEWT = Stachowski 2019
 TMEN = Doerfer 1967

- 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.
 Doerfer G. 1967. *Türkische und mongolische Elemente im Neupersischen*. Wiesbaden: Franz Steiner.
 Erdal M. 1998. Old Turkic. – Johanson L., Csató É.Á. (eds.). *The Turkic languages*. London, New York: Routledge: 138–157.

- Erdal M. 2022. On difficulties encountered in etymologizing the Turkish lexicon. – *International Journal of Eurasian Linguistics* 4: 75–83.
- Grønbech V. 1902. *Forstudier til tyrkisk lydhistorie*. København: Lehmann & Stage.
- Güzel F. 2019. *Türkiye Türkçesi ağızlarında yapım ekleri*. Ankara: Türk Dil Kurumu.
- Hirik E. 2012. Review of Stachowski 2011. – *Karadeniz Araştırmaları* 34: 184–189.
- Levickij V.V. 2010. *Ėtimologičeskij slovař germanskich jazykov – An etymological dictionary of Germanic languages*. Vinnica: Nova Knyha.
- Miller R.A. 1991. How dead is the Altaic Hypothesis? – Brendemoen B. (ed.). *Altaica Osloensia*. Oslo: Universitetsforlaget: 223–237.
- Nadeljaev V.M. 1969. *Drevnetjurkskij slovař*. Leningrad: Nauka.
- Öztürk S. 2012. Review of Stachowski 2011. – *Türkbilig* 23: 153–160.
- Sevortjan È.V. 1974. *Ėtimologičeskij slovař tjurkskich jazykov*. [vol. 1]: *Obščetjurkskie i mež-tjurkskie osnovy na glasnye*. Moskva: Nauka.
- Şirin User H. 2004. Türkçede “para” ve “para birimi”. – *Türk Dili Araştırmaları Yıllığı – Belleten* 2004: 129–151.
- Stachowski M. 2011. *Etimoloji*. Ankara: Türk Kültürünü Araştırma Enstitüsü.
- Stachowski M. 2019. *Kurzgefaßtes etymologisches Wörterbuch der türkischen Sprache*. Kraków: Księgarnia Akademicka.
- Tietze A. 2002. *Tarihi ve etimolojik Türkiye Türkçesi lughatı. Sprachgeschichtliches und etymologisches Wörterbuch des Türkei-Türkischen*. [vol. 1]. İstanbul: Simurg, Wien: Österreichische Akademie der Wissenschaften.
- Wilkens J. 2021. *Handwörterbuch des Altuigurischen. Altuigurisch – Deutsch – Türkisch*. Göttingen: Universitätsverlag.
- Yıldız H. 2011. Review of Stachowski 2011. – *Dil Araştırmaları* 9: 179–194.