FOLIA ORIENTALIA
VOL. 34 1998
PL ISSN 0015-5675

MAREK STACHOWSKI

A MINIMAL PROBABILISTIC DEVELOPMENT MODEL OF PROTO-TURKIC *E*-TYPE VOWELS

1. As far as *E*-type vowels in Tkc. languages are concerned, one has to distinguish between two groups of languages: (a) those with an open *ä* alone and (b) those with an open *ä* and some other *E*-type vowel(s). It is not always possible to establish precisely how many *E*-type vowels in group (b) possess the status of a phoneme, the main reason for which being the often inaccurate description of the Tkc. vowel systems we have at our disposal. In any case, there can be no doubt that some Tkc. languages (as e.g. Kirg. and Tuv.) have only one *E*-type phoneme (i.e. an open *ä*) while others have additionally another phonemic *E* vowel, namely a closed *e*.

The problem of one or more than one *E*-type vowels is one of the oldest questions of Turkic linguistics. It was already Vilhelm Thomsen (1916: passim) who first assumed the existence of *e*, as opposed to *ä*, in Orkhon Turkic, and then was able to prove his assumption after having identified a Yenissei Turkic rune, which was missing from the Orkhon runic alpha-

bet, as a letter for *e*. The problem has, as can be seen, a long history, and it is not our goal here to make a summary of it.

The purpose of our modest note is rather to establish a minimal probabilistic model of the development of *E*-type vowels, i.e. to find out how many *E* phonemes must be reconstructed in the PT vocalic system. Moreover, some other remarks will be made which do not concern so much the PT vocalism only but rather a more general problem of Tkc. historical phonology, i.e. the mutual correlation of comparative data.

2. Our starting point is a simple scheme:

[2.1] PT had **ä* only

[2.2] PT had **e* only

[2.3] PT had both **ä* and **e*

Of the three possibilities, [2.2] can be left aside as most unlikely because no Tkc. language has exclusively an *e*. If we now combine the problem of *E*-type vowels with that of vocalic quantity,[1] our scheme will become somewhat more complex:

[2.4] Four PT vowels: **ä*, **ā̆*, **e*, **ē*

[2.5] Two PT vowels: **ā̆*, **e*

[2.6] Two PT vowels: **ä*, **ā̆*

[2.7] Two PT vowels: **ä*, **ē*

Some possibilities have been left aside here: (a) PT **e*, **ē*; (b) PT **ä*, **e*; (c) PT **ä*, **ā̆*, **e*; (d) PT **ä*, **ā̆*, **ē*; (e) PT **ä*, **e*, **ē*, (f) PT **ā̆*, **e*, **ē*. The reason for excluding (a) is the same as that of excluding [2.2]: there exists no Tkc. language

without an open *ä*. The (b) variant cannot be taken into account for it shows no long vowels. More problematic are (c) - (f) cases. The problem is somewhat broader and it involves two other questions. The first one is whether the postulated **ë* vowel which stood between **a* and **y* actually was a phoneme. If it were so, we really could postulate a three-fold opposition both for velar and palatal vowels: **a* : **ë* : **y* and **ä* : **e* : **i*. But in reality, nothing seems to point to the phonemic status of **ë* (for more details see Stachowski 1993: 23-27). If there was only a two-fold opposition of velar phonemes in PT (**a* : **y*), another question arises: may we reconstruct an asymmetrical vowel system, i.e. **a* : **y* and **ä* : **e* : **i*? The possibility of such a system cannot be wholly excluded but it is most unlikely for there are no other three-fold vowel oppositions in Tkc. Nor does there exist a Tkc. language with **a* : **ë* : **y* as phonemes. In this situation, the acceptance of (c) - (f) were only possible if we could give positive answers to two questions, whereas the first answer is rather negative and the second may be either positive or negative. For the reliability of the probabilistic model, it is, as it seems, better to leave the three-fold opposition aside and to see if the model can work with two original *E*-type phonemes only.

Another limitation of the present study concerns positional changes like **jä* > *je* (Doerfer 1994: 112, 123) which do actually occur on the phonetic level but need not necessarily cause a restructuring of the phonological system and therefore can be left aside when constructing a minimal probabilistic model.

The variant [2.5] is hardly a real possibility. It is generally long open vowels which tend to be narrowed (the cause of which is a purely physiological tendency to minimize the articulatory effort - it is easier to pronounce a long vowel without holding the mouth wide open all the time; cf. also Doerfer 1994: 110: "Auch im Deutschen wird übrigens "Träne", eigentlich /trānə/, meist /trēnə/ gesprochen"). Consequently, it would rather be expected that a long vowel is closed (= narrow) and a short one is open. This is exactly the case in [2.7]. The possibilities [2.6] and [2.7] can be viewed as two chronological stages of one and the same narrowing process:[2] [2.6] + [*ā̈ > *ē] → [2.7]. But the narrowing of an open vowel needs not necessarily oust the original vowel (i.e. *ā̈ > *ē) which, on the contrary, can persist alongside with the new vowel that was produced out of it, i.e.

[2.8] PT *ä : *ā̈ > PT *ä : *ā̈ ~ *ē

3. The next problem is if there ever existed an original short closed *e. It was N. Poppe (1925: 413f.) who proposed Chuv. and Az. as the key-languages in this matter by laying down the following rules:

[3.1] PT *ä > Chuv. *a*, Az. *ä*

[3.2] PT *e > Chuv. *i*, Az. *e*

Both rules have been repeated many times in the literature and they seem to give us a really practical key for our reconstructions. In reality, however, the use of the key is not that easy for it ignores both Trkm. and Yak. phonetic variants, cf. the

following equations:

[3.3a] Chuv. *sas* 'voice' < **säs* > Trkm. Az. *säs* id.

but **[3.3b]** Chuv. *kaś-* 'to go through' < **käč-* > Trkm. *gäč-*, Az. *keč-* id.

and

[3.4a] Chuv. *ir(-ĕl)-* 'to melt [intr.]' < **er-* > Yak. *ir-* id.

but **[3.4b]** Chuv. *śil* 'wind' < **jel* > Yak. *säl* id.[3]

As can be seen, the reconstruction based on Chuv. data cannot be readily correlated with the data of other Tkc. languages. The use of the Az. word in [3.3b] as a starting point would lead us to **keč-*, not to **käč-*; similarly, Yak. *säl* in [3.4b] suggests **jäl* rather than **jel*. It goes without saying that reconstructions must not depend on which language was chosen by the respective scholar as his starting point.[4] An ambiguous reconstruction like **sEs*, **kEč-*, **Er-* or **jEl-* (with *E* = *e* or *ä*) does not solve the problem but it is, in fact, more honest than choosing one language in a relatively arbitrary way and forcing reconstructions dependent on its vowels upon all the other languages.

Moreover, Poppe's rule cannot be used in each case, e.g. not for

[3.5] Chuv. *üt* 'flesh; meat' (= Tkc. *ät* id.),

which shows neither *a* nor *i*.

Both problems (the lack of correlation[5] with other Tkc. languages; vowels other than *a* or *i*) can be readily exemplified by the following instructive examples:

[3.6] Chuv. *ĕm-* 'to suck' = Yak. *ăm-* id.

[3.7] Chuv. *śĕr* 'earth, soil' = Yak. *sir* id.

The Chuv. words in both examples have an *ĕ* that is neither *a* nor *i*, and their Yak. equivalents show two different reflexes. In other words: if one wished to start from Chuv. one could not achieve any result at all; if one started from Yak. one would have to reconstruct two protoforms with two different vowels which would have one and the same reflex in Chuv. but the reflex were neither *a* nor *i*.

The assumption of a PT **e* does not seem to make the problem easier. It is therefore preferable to see if the development model of *E*-type vowels could work without an original short **e*.

4. Let us take [2.6] as our starting point. It has been shown how a secondary **ē* could have evolved, see [2.8]. The next step is a neutralization of the vowel quantity opposition in some PT dialects (**ā̆* > **ă*; **ē* > **e*) and still a further one is the narrowing tendency which led to a partial merger of **e*, **ē* with **i*, **ī*. All that can be shown in a four-stages scheme:

[4.1] Stage 1: PT **ă* : **ā̆*

As a result of a narrowing process of vowels, the phoneme */ā̆/* produced an additional allophone **[ē]*, i.e.

[4.2] Stage 2: dialect A: **ă* : **ā̆* ~ **ē* (< **ā̆*)

dialect B: **ă* : **ē* (< **ā̆*)

The shortening of vowels in the most vernaculars divided the Tkc. linguistic world into two dialects (~ dialectal groups):

[4.3] Stage 3: dialect C: **ä : *ā̈ ~ *ē*
dialect D: **ä (< *ä, *ā̈) : *ė*[6]

The partial narrowing of **ē > *ī*, **e > *i* produced words with a secondary **ī* and **i* which corresponded to **ā̈ ~ *ē* and **ä ~ *e* in other idioms:

[4.4] Stage 4: dialect E: **ä : *ā̈ ~ *ē ~ *ī*
dialect F: **ä : *e ~ *i*

If a secret power had separated the Tkc. idioms from each other after Stage 4 had been reached we would probably have had a much more consequent and clear picture today. But there was no secret power. The Tkc. vernaculars continued to be used in neighbouring territories and mutual influences and borrowings made the picture fairly blurred. That is why we cannot view our hypothetical dialects E and F as direct successors of A and B, respectively, but rather as new configurations of Tkc. dialects (cf. note 6).

Very interesting is the equation: Yak. *iä* = Az. *e* (Thomsen K 1957: 152f.), as in Yak. *biäs* 'five' = Az. *beš* id., Yak. *diä-* 'to say' = Az. *de-* id. K. Thomsen tried to explain the situation by the *ė > ā̈* change (opening + lengthening) which seems quite improbable. We are rather inclined to postulate a chronological dialect shift: B → D. Similarly, Trkm. gives examples both for *i < *ā̈* (Trkm. *di-* 'to say' = Yak. *diä-* id.) and for *ā̈ < *ā̈* (Trkm. *bāš* 'five' = Yak. *biäs* id.), the former reflecting the B → D → F shift, the latter the A → C → E one. But we cannot exclude a more complex development line, either, like A → C → D → F or A → C → E → F (both for Trkm. *di-*) and so on.

Cf. also the following opinion: "Im großen und ganzen sind die Vertretungen in den Türksprachen ziemlich inkonsequent, was in vielen Fällen eine Folge der Sprachvermischung und gegenseitiger Entlehnung sein könnte" (Poppe 1960: 105). Unfortunately, N. Poppe does not offer any concrete solution to the problem.

Another problem is a secondary redistribution of *E* vowels like in Uyg. (*ä* in closed syllables [e.g. *käč* 'late'] vs. *e* in open syllables [e.g. *kečä* 'evening'] but cf. also Sven Hedin's records: *ördäk* ~ *ördek* 'wild duck, duck', *mähällä* ~ *mehelle* [~ *mehalle*] 'district, quarter', etc., and G. Jarring's (1997: XI) comments: "The amplitude ɛ [= *ä*] ~ *e* is quite normal in the spoken language") or partially in modern Turkish (cf. "Ein *e* der ersten Silbe, das von einem [...] *i* der zweiten Silbe nur durch einen (kurzen) Konsonanten getrennt ist, hat immer geschlossene Qualität" [Collinder 1939: 13]).

5. It cannot possibly be a pure coincidence that the Chuv. reflex of original **ä* and **ā̈* always is an *a*, whereas **ē* should have resulted in Chuv. *i*, *ĕ* or *a*, e.g. **ēki* 'two' > Chuv. *ikkĕ* id. (cf. § 6), **ēkiŕ* 'twin' > Chuv. *jĕkĕr* id. (cf. § 6), **bēr-* 'to give' > Chuv. *par-* id. (Ceylan 1997: 159f.); note also the irregularity of Yak. correspondences: Chuv. *ikkĕ* 'two' = Yak. *ikki* id., Chuv. *par-* 'to give' = Yak. *biär-* id. It appears to us impossible to accept several reflexes of PT **ē* without having formulated concrete phonetic conditions which entailed a different development of the vowel. If such a formulation is not

possible, one should rather accept a dialectal mix-up (and wait for a future, possibly phonetic solution).

Word equations like:

[5.1] **bēl'* 'five' > Volga-Bulg. *bijäl(-im)* > Chuv. *pil(-lĕk)* = Yak. *biäs* id.

[5.2] **jēti* 'seven' > Volga-Bulg. *ǧijäti* > Chuv. *śiччĕ* = Yak. *sättä* id.

display a zigzag development (which is in principle an uncertain idea): **ē* > *ijä* > *i* and it makes no sense to correlate Chuv. and Yak. reflexes.

6. The fact that the Yenissei runic script made use of a special rune for closed *e* and the Orkhon script did not, seems to indicate that the rune for *e* probably was a new invention that had not yet spread over the whole territory of Old Tkc. vernaculars. It is difficult to say what it means to the chronology of the phonetic change. The closed *e* was probably no full phoneme of equal status in the system, at that time. Otherwise, it would be expected that the invention of an *e* rune would immediately spread throughout the Tkc. idioms. But it can at least be assumed that the narrowing process has been in progress in the seventh/eighth century for it left its traces in Runic inscriptions, cf. "**ā* ist recht früh > /ē/ geworden; man darf davon ausgehen, daß für alle Inschriften /ē/ gilt [...]" (Doerfer 1994: 112). If Erdal (1993: 145) is right in his assumption that the *ä* ~ *i* alternation (as in Bulg. *äki* 'two' vs. *ikiz* 'twin') is a genuine phenomenon rather than a reflex of an

old /e/, Bulg. data cannot be easily used for chronological purposes. The dialect D (Stage 3) could only have evolved after the shortening of vowels had started. In Doerfer 1975/76: 131 the shortening process is finished towards the end of the 12th century and this period can be put as the initial phase of the development process of the dialect D so that its end phase must have fallen into the 13th century. In other words: the most important changes (Stage 2 and 3) seem to have covered the period from the 7th till the 13th century. Other far-reaching inferences that could have been drawn from the outlined reconstruction of the development process seem to us rather risky.

7. Our analysis has, as we hope, shown that Stage 1 of the minimal probabilistic model of the development of PT *E*-type vowels consists of **ä* and **ā*. The assumption that the short narrow **e* existed already in Early PT as an independent phoneme is not necessary[7] for reconstruction; moreover, it leads to serious problems in correlating comparative data from different languages.

Another (parallel) possibility of the development of *e* is that *e* originally was an allophone of both **ä* and **i* which fits quite well to the well-known problem of (Németh:) **a*$_2$ = (Doerfer:) **ë* (see above § 2). This possibility does not, however, change the proposed probabilistic model because an allophone (i.e. [e]) must have been secondary in respect to the

phoneme (i.e. /ä/, /ā̈/, /i/ and /ī/) and, therefore, it cannot be readily used in the reconstruction of a minimal model.

* * *

I would especially like to thank E. Helimski, from the discussion with whom I have significantly benefited.

Notes

1. I find myself compelled to refrain from taking the so called diphthongic vowels of Khalaj into consideration because their importance for the Tkc. vocalic reconstruction is questionable, cf. the following two opinions, both belonging to G. Doerfer: [1] "Proto-Turkic originally did not have two "quantities" but three", Doerfer 1971: 234 (in capital letters!); it is somewhat astonishing to see that elsewhere, the same author writes on the same issue as follows: "Auf eine definitive Behauptung über die Quantitätsverhältnisse des Urtürkischen hatte ich verzichtet" (Doerfer 1988a: 23) and then, still in the same year, he expresses a totally opposite opinion: [2] "(...) für das Urtü. dürfen wir also zwei Quantitäten ansetzen" (Doerfer 1988b: 11); "Im Urtü. gab es nur zwei Vokalquantitäten" (ibid. 12).

2. K. Thomsen (1957: 153) views it differently but we can hardly accept his assumption that [2.7] was the original situation because his arguments are rather weak (cf. note 4) and his solution does not clear the problem why the long equivalent of **o* was an open **ō* whereas the long counterpart of **ä* was not an open **ǟ* but a narrow **ē*. Moreover, the idea that [2.7] was original and **ǟ* was secondary would lead us to the conclusion that **ē* > *ǟ* which contradicts to the data of typology.

3. Cf. also Yak. *kim* 'who?' ~ Chuv. *kam* id.; Yak. *ilī* [~ *älī*] 'hand' ~ Chuv. *ală* id.; Yak. *dirıŋ* 'deep' ~ Chuv. *tarăn* id.; Yak. *tir-it-* 'to sweat' ~ Chuv. *tar* 'sweat' (Räsänen 1949: 89). Besides, note that the Yak. reflex of OT *ēš* 'duty; work' is *iäs* (Doerfer 1993: 72, 73; Doerfer 1994: 108).

4. Cf. K. Thomsen's (1957: 153) opinion on **ē*: "(...) the best criterion is Azerbaijanī *ė* (alternation *ä/i* in the other dialects) together with clear evidence of length"; by the way, this statement does not fully correspond to what the same author says some lines earlier, but still on the same page of his article: "Common Turkish had (...) *ä* which in palatal surroundings (after or before *j* and before *i*) facultatively became *i* and in Azerbaijanī *ė* (...)". A wholly opposite opinion was expressed by V. Grønbech (1902/79: 127): "[...] there is no reason to distinguish Orkh. *in-* 'to climb down' [...] = Koib. *en-*, Chuv. *an-*, cf. Yak. *äniä*

'mountain slope', Viguier writes *en-*, Kúnos *in-*; whatever the reason may be for this modern *i*, it certainly does not point to any old distinction in the vowel." It is incomprehensible, indeed, why the *ä* ~ *i* alternation should have pointed to an original opposition while the *ö* ~ *ü* (as in Oyr. *öč* 'Rache' ~ Khak. *üs* id., Radloff 1882: 86) and the *o* ~ *u* alternation (as in Oyr. *soŋda* 'nachher' ~ Tel. *sunda* id., ibid.) should not, cf. for instance Erdal 1996: 76, 79–81, where no vowel like a closed **o* or an open **u* (that is one standing between *o* and *u*) has been suggested for word pairs with the *o* ~ *u* alternation.

5. The lack of correlation does not concern Chuv. and Yak. data only, cf. Räsänen (1949: 90): "Diese [= Turkish *e* words] aber stimmen nicht gut mit čuv.-az.-Vergleichungen überein, und ausserdem scheint in dieser Hinsicht ein grosser Wirrwarr im osm. [= Osm.] vorzukommen wie auch in den anderen Türksprachen.".

6. As can be seen, dialect D needs not have been a simple continuation of B. One can easily imagine that the shortening of vowels divided the dialect area A in two parts: A_1 and A_2 and that long vowels in A_1 were preserved, whereas those in A_2 became short, so that A_2 merged with B and both groups, A_2 and B evolved into D.

7. Nothing more can be expected from a probabilistic model. It is not in a position to exclude a possibility entirely; the only goal of applying minimal probabilistic models is to establish what is realistic and absolutely necessary and to separate it from what is possible, albeit not absolutely necessary.

Abbreviations

Az. = Azerbaijanian; **Bulg.** = Bulgar; **Chuv.** = Chuvash; **Kirg.** = Kirgiz; **PT** = Proto-Turkic; **Tkc.** = Turkic; **Trkm.** = Turkmen; **Tuv.** = Tuvinian; **Uyg.** = Uygur; **Yak.** = Yakut.

References

Ceylan 1997 = Ceylan, E.: *Çuvaşça çok zamanlı ses bilgisi*, Ankara.

Collinder 1939 = Collinder, B.: *Reichstürkische Lautstudien*, Uppsala.

Doerfer 1971 = Doerfer, G.: *Khalaj materials*, Bloomington.

Doerfer 1975/76 = Doerfer, G.: Das Vorosmanische (Die Entwicklung der oghusischen Sprachen

von den Orchoninschrfiten bis zu Sultan Veled). - *Türk Dili Araştırmaları Yıllığı. Belleten*: 81-131.

Doerfer 1988a = Doerfer, G.: Zur Quantität der Vokale des Chaladsch. - *Wiener Zeitschrift für die Kunde des Morgenlandes* 78: 23-73.

Doerfer 1988b = Doerfer, G.: *Grammatik des Chaladsch*, Wiesbaden.

Doerfer 1993 = Doerfer, G.: Zu alttürkisch *Is²g ~ s²g küčg b²Ir²-* 'dem Herrscher gegenüber seine Pflicht erfüllen, ihn unterstützen'. - Kellner-Heinkele, B. (ed.): *Altaica Berolinensia*, Wiesbaden: 69-74.

Doerfer 1994 = Doerfer, G.: Zu inschrifttürkisch *ē/e*. - *Ural-Altaische Jahrbücher. Neue Folge* 13: 108-132.

Erdal 1993 = Erdal, M.: *Die Sprache der wolgabolgarischen Inschriften*, Wiesbaden.

Erdal 1996 = Erdal, M.: Zum alttürkischen Vokalsystem. - Emmerick, R. E. et al. (eds): *Turfan, Khotan und Dunhuang*, Berlin: 67-82.

Grønbech 1902/79 = Grønbech, V.: *Preliminary studies in Turkic historical phonology*, Bloomington 1979 [translated from the Danish by John R. Krueger; original title: *Forstudier til tyrkisk lydhistorie*, København 1902].

Jarring 1997 = Jarring, G.: *Central Asian Turkic place-names (Lop Nor and Tarim area). An attempt at classification and explanation based on Sven Hedin's diaries and published works*, Stockholm 1997.

Poppe 1925 = Poppe, N.: Türkisch-tschuwassische vergleichende Studien. - *Islamica* 1/4: 409-427.

Poppe 1960 = Poppe, N.: *Vergleichende Grammatik der altaischen Sprachen*, Teil 1: *Vergleichende Lautlehre*, Wiesbaden.

Radloff 1882 = Radloff, W.: *Phonetik der nördlichen Türksprachen*, Leipzig.

Räsänen 1949 = Räsänen, M.: *Materialien zur Lautgeschichte der türkischen Sprachen*, Helsinki.

Stachowski 1993 = Stachowski, M.: *Geschichte des jakutischen Vokalismus*, Kraków.

Thomsen K. 1957 = Thomsen, K.: The closed "e" in Turkish. - *Acta Orientalia* (København) 22/3-4: 150-153.

Thomsen V. 1916 = Thomsen, V.: Une lettre méconnue des inscriptions de l'Iénissei. - *Journal de la Société Finno-Ougrienne* 30: 1-9.