

## Hypercorrect orthographic forms in the Pagan Oğuz-nāmā -- a phonetic analysis<sup>1</sup>

The Pagan Oğuz-nāmā (MS, Radloff 1891, Nour 1928, Pelliot 1930 (1995), Bang-Rachmati 1932 (1936), Ščerbak 1959) is a Middle-Turkic text written in Uygur script, and consist of 42 pages with 9 lines on each page. Pelliot's (1930) view, that it has been written around the beginning of the 14th century is accepted by the scholars who dealt with this text later. This view also occurs in Clauson's dictionary (ED 39.§, XXIII). Pelliot also pointed out some phonetic and orthographic features of the text, and in his final conclusion proposed that the text should have been an original which is lost, and the MS we have in hand is a later copy from the 15th century, made by a scribe who spoke some Kipchak dialect.

The Pagan Oğuz-nāmā (in the following, PON) so far has been proved to be problematic to classified among Turkic monuments and it is not a surprise that the latest edition of the text is from the end the fifties of the last century, due to the PON's unusual orthographic and linguistic features. The orthography and the language of a text are in intimate contact with each other, and it is of crucial necessity to understand this relation in order to describe the language of the PON. The way to this goal – considering the fact that the PON is a *written* monument of a *once living, but already dead language* -- leads through the examination of the orthography, which is the basic source of the phonetics and phonology of this language.

The orthography of the PON has the following general features:

1. There are no solid rules to mark the quality of a given sound precisely. This leads to fluctuating depiction of the individual lexemes, which fortunately allows us to draw conclusions about certain phonetic phenomena of the dialect of the text. In other words, the orthography of the text is rather phonetic than phonologic.

2. There is no orthographic distinction between [q] and [ɣ] (marked originally with <q> and <ḡ> respectively in uygur script). Both allophones can be marked with either <q> or <ḡ>. In the following, I will refer to both as <q>, where there is no relevance of the presence or absence of the diacritic dots.

3. There is no orthographic distinction between /t/ and /d/ both are marked with <d>

4. The does not make distinction between /s/ and /z/. The orthography uses two graphemes to mark these sounds <s> and <-z> respectively, but their distinction is almost distributional, <-z> (homoformous with <-ʻ>) always occur on the end of a grapheme sequence, marking both /s/ and /z/, while <s> can occur only on the initial or internal part of a grapheme sequence. There are a couple of exceptions from this rule, in these cases the lower stalk of the <s> is missing, thus the graph looks very similar to <-l>.

---

<sup>1</sup> The present paper has been prepared in the framework of the *Turkological Research Group of Hungarian Academy of Sciences and University of Szeged*.

The publication is supported by the European Union and co-funded by the European Social Fund.

Project title: "Broadening the knowledge base and supporting the long term professional sustainability of the Research University Centre of Excellence at the University of Szeged by ensuring the rising generation of excellent scientists."

Project number: TÁMOP-4.2.2/B-10/1-2010-0012

5. There is no orthographic distinction between labial vowels. All labial vowels are marked with <'w> (or even simply with <w>) in word-initial position and with <w> in word-internal or word-final position. The original method in uigur script was <'w> for word-initial /o/ and /u/, and <w> for non word-initial /o/ and /u/, and <'wy> for word-initial /ö/ and /ü/, <wy> for /ö/ and /ü/ in first syllable, and <w> for the remaining positions. This simplified marking method may show influence of arabic script, which gradually squeezed and replaced uigur script among the Turks until the first half of the 15th century.

6. The orthography of the PON seems to show influence of Written Mongolian orthography in the way of marking primary or secondary long /a:/ sounds by the grapheme-sequence <'q'>. Several words occur in the text with this hypercorrect orthography, which are already registered in the previous editions of the PON, but no exhaustive examination has been made covering the whole text, although on a second glance, the keen eye may find other examples, but written with <'qw>, <'qy>, etc.

The aim of the present paper is to introduce these words with hypercorrect orthography, and to attempt to determine the quality of the sounds marked by <VqV>.

In order to reach our goal, it is subservient to introduce briefly the history of Mongolic sounds marked by <VqV> in Written Mongolian.

The parallels of Written Mongolian <'q'>, <'qw>, <'qy> etc. grapheme sequences in modern Mongolic languages are secondary long vowels or diphthongs due to the result of omitting of /g/ in intervocalic position. The orthography of the Chinese transcripton (end of 14th century) of the Secret History of Mongols (originally from 1227) allows the conclusion that in contemporary spoken Mongolian the process of loss of the intervocalic /g/ had been already started, and has a so-called 'hiatus' in intervocalic position. Pre-classic and Classic written Mongolian however, preserved the marking of intervocalic /g/. Scholars still argue about the actual quality of this etymologic /g/ in 14th century Mongolian language, but the detailed argument of the topic is not the subject of the present paper. The two most recent opinions are the following:

According to Miller's (2002) very detailed analysis, that the sound in question had been a laryngeal or uvular weak spirant [h̥] which developed from a \*[ɣ], and this [ɣ] is marked in by <k> or <q> in Written Mongolian. Janhunen derives the intervocalic „hiatus” from a „Pre-Proto-Mongol” \*p which developed into \*χ and the language of the Secret History show this sound as a laryngeal [h].

Now we can turn to the data of the PON. In every example I gave the locus of the given words in the manuscript as X/Y where X stands for the number of the page, and Y for the number of the line on the page. This also means that every example I give are word instances and not lexemes, however, many of them occur only a single time in the text. Consider the following examples:

- (1a) <q'q'r> 27/9                      ~ OT *qa:r* 'snow' ED 461  
 (1b) <q'q'r-l'q> 28/3                    'proper name, Qarluq' < OT *qār* +IXg<sup>2</sup>

---

<sup>2</sup> The context provides a folk etymology of the Turkic tribal name. In the text the the protagonist gives the name because the *beg* is covered by snow, this verifies that the word formative element of the word corresponds to Old Turkic +IXg, although it is homographous with +IXK.

- (2) <q'q'd'r> 31/3 ~ OT *qatir* 'mule' ED 604, قاطر Z 682a
- (3) <š'q'm> 33/3 ~ ar. *ša:m* شام 'Syrien, Damaskus' Z 536a
- (4) <d'q'm> 28/6 ~ OT *ta:m* 'wall' ED 502

As we can see, the the spelling of <'q'> is unmotivated in the sense that none of the words contains an etymological /g/. In the examples (1a), (1b) and (4) the vowel of the word stems goes back to etymological Old Turkic long /a:/. The etymon of example (3) goes back to an Arabic word which also contains a long /a:/. Example (2) etymologically contains no long vowel, but the spelling of the word in Arabic script contains the grapheme < ʾ >, which originally marks long /a:/. The Arabic spelling of the word may had influenced the pronunciation of the word.

As we have seen, all the examples contain <'q'> spelling, which suggests that they contain that they had been pronounced with long /a:/ at the time of the born of the text. The spelling of the words suggests that the scribe had competence in written Mongolian. The fact that these spellings , with the exception of (2), contain etymologic long vowel, also suggest that the spoken Mongolian variety known by the scribe had contained already secondary long /a:/, which developed from the disyllabic /aga/. Since the 14th century copy of the Secret History still had these disyllabic sequences, the rising of the text must show a later stage of this development, or a dialect which had been more progressive in this aspect. Now let us turn to next example:

- (5) <d'qwr'q> 13/6 ~ OT *taβraq* 'speed, hurry, quick' ED 443

The Old Turkic word had never contained etymological /g/. The corresponding sound of OT /aβ/ is spelled by <'qw>. I assume that the instance occurring in the PON contains a diphthong with a labial element (marked by <w>), which developed from the phoneme sequence /aβ/. I will return to its closer quality later. Another case is visible at example (6):

- (6) <d'rl'qw-syz> 29/8 ~ OT *tarlağ*, *tarla* ED 546 ~ *tariğlağ* 'a cultivated field' ED 541 +SXz

In this case the instance is spelled by <w> again, however, the etymon of the word lacks a labial element. The word can only be derived from *tarlağ*.<sup>3</sup> The phoneme sequence /ağ/ is spelled also with <'qw>, thus its pronunciation must had coincided with that of /aβ/. This phenomenon, which typical of Kipchak languages, occurs already in the Italian part of the Codex Cumanicus from the 13th century (the word in question occurs there as *tarlov*, ED 541).

---

<sup>3</sup> It would be possible to subtract a word formative –*GUSIz* element (with the opposite meaning of –*GUIUK*, however, a verb stem *\*tarla-* < *\*tariğla-* could not be dated, only the derivated form *tarlağ* < *tariğlağ* (ED 541)

Another word may be added to our list so far, which seems to be Mongolian origin, although somewhat problematic:

(7) chag. قوریا *kurja* 'Mauer, Hütte (Z 716c) ← mong. *qoriy-a(n)* 'enclosure, camp' (L 967)

(7a) <*qwryq*'n> 12/3

(7b) <*qwryq*'n> 14/9

(7c) <*qwryq*'n> 15/3

(7d) <*qwr*'q'n> 17/4

(7e) <*qwryq*'n>/<*qwr*'q'n> 29/7

One problem with this word is of semantic nature. Considering the contexts in which the word occur, it means '(Oğuz Qağan's) tent' and always not 'camp' or 'enclosure (for animals)'. The Chagatay word does have the meaning 'hut (of poor quality)' but not a 'tent of the ruler'. The word قوریا seems to have the secondary meaning from the (reborrowed?) Mongolian word. The other problem is that if the scribe had competence in (Written) Mongolian, why didn't he use the Written Mongolian form of this word? Instead, he used again the spellings <yq'> and <'q'> for the phoneme sequence /iya/, which, at least on the Mongolian side did not contain /g/ or /k/. In any case, based on the examples cited above, the pronunciation of the word should have contain a diphtong (<yq'> spellings) or long vowel (<'q'> spelling(s)).

Until this point, we saw hypercorrect examples which never contained etymological /g/ but are spelled so with Mongolian orthography, thus allowing the assumption that they contained diphtong or long vowel. Now let us turn to another group of data, which seems to fall under the same orthographic rule, but the reason of their spelling is different. These words the syllable-structure (C)VgVC, and fall under the phonotactic rule that if they are take suffix with initial consonant, the second-syllable vowel becomes omitted. In these cases however, the spelling of the words in question does mark second-syllable vowel:

- (8) *ayiz* +Px3(+Cx) 'his/her mouth' > *ayzi*

(8a) <'qysy>/<'q'sy> 1/6

(8b) <q̇'y<sup>z</sup>-wm-q'> 13/1 (sic)

(8c) <'q'z-wm-q'> 13/4

(8d) <'q'z-y-q'> 14/5

(8e) <'q'sy-q'> 30/8

Example (8b) is misspelled, the diacritic dots are put next to the first two <'>-s, not the second two ones. In (8a) it is not completely clear that whether a <y> or an <'> to be read in the second syllable (The only difference between the two graphemes is their length). In these cases, the data can be analysed in two ways: 1. The morphophonetic rule mentioned above is not valid; 2. The second syllable vowel is a part of an <'q'> or <'qy> sequence again, thus it is to be read together as a secondary long vowel or diphtong. Based mainly on (8c), (8d) and (8e), and on (9) cited below, we may assume that we confront the latter case.

(9) OT *yayir* 'a saddle-gall' > 'shoulders' (ED 905)

(9a) <yq̣'ry> 2/4 occurring two times in the same line

The ED cites the corresponding part of the PON as *yayri kiš yayri teg* 'his sholders like a sable's', however, the spelling of the word is as in (9a), as the diacritic dots occurs immediately after the word-initial <y>. If Clauson's reading (based on Bang's) is correct, and the word *yayir* is to be seen here, and not the Sogdian loanword <yγ'r> *iγar* '(„probably") strong' (ED 89), in the position of a noun (meaning 'strength'), then we can see the same sequence as in (7a), (7b) and (7c). For the lack of the initial <'> denoting word-initial vowel see also (8) <'q'z> instead of the expected <'q'z>, and any words in the text with /a-/ and some of /i-/. Ultimately, the second syllable <'> in (9a) instead of <y> suggests that the grapheme sequence <yq̣'> in (9a) and <'q'>/<'qy> in (8) should be read together as secondary long vowel or diphtong respectively, and should not be considered as the invalidity of omitting the second-syllable vowel before vowel-initial suffix.

Further data can be observed with similar morphophonetic environment with labial vowels:

(10) *kögüz* 'chest, breast' (ED 714) +Px3 > *kögzü* 'his or her breasts'

(10a) <kwkwzwndwn> 1/9

(10b) <kwkwzw> 2/4

(10c) <kwkwzw> 2/5

(11) *oyul* 'offspring, child' (ED 83) +Px3 > *oyli* 'his/her son'

(11a) <'wqwl-wn> 20/3

(11b) <'wqwl-y> 21/1

The next group of data shows that the diphtong-marking <VqV> is can overarch morpheme-boundaries:

(12) chag. چیرا، چیرلی čira, čiray 'Gesicht, Wange' (Z 378a) < Mong. čiray (L 191)+Px3

(12a) <č'r'qy> 1/5

(12b) <č'r'qy> 34/3

(13) <swqy> 19/4 ~ OT *suβ* 'water' ED 783 + Px3

Again, in (12) and (13) there is no etymological /g/, <'qy> and <wqy> marks word-final diphtongs to which a vowel-initial suffix had been added.

Some instances of deverbal nominal formative –GU points that the pronunciation of this formative already had diphtongic value:

(14) OT *ay-* 'to remember, call to mind' (ED 168) –GU +Px3<sup>4</sup>

(14a) <'nk'qw-sw> 1/1

(14b) <'nkqw-sy> 5/8

(14c) <'nkwqw-sw> 6/3

(15) <'wq'-qw-luq> 35/8 ~ *uq-* 'to understand something' (ED) 77 –GU +IXg

The 'second syllable' vowel graphs in (14a), (14c) and in (15) could be also analysed as anaptyctic sounds, but in the mirror of the above train of thought, these data also belong here, the <'qw> and <wqw> grapheme-sequences should mark a diphthong or secondary long-vowel respectively.

In the cases of the above data, the verbs stems end in a consonant, but in the case of *ay-* two of the three instances an additional vowel-marking grapheme occurs, similar to the only instance of *uq-*. This could be analysed as an anaptyctic sound (to the existence of anaptyctics see <*kwrwk-lwg*> ~ *körk+lXg*, <*qwrwq-m'z*> ~ *qorq-mAz*, <*d'r'ddy*> ~ *tart-DX* etc., (the rule could be formalised as [ (C)Vr°CC\_ ← (C)VrC+C\_ ]), but here the phonetic environment is different, and could be formalised as [(CVG°GW ← (C)VG+GW] which in the end provides a /VgV/ sequence, marked as <'qw> and <wqw> respectively, which could be (and I think it must be) read as diphthong or secondary long vowel, thus it ends up in the same result.

Another reason that the anaptyctic sounds are improbable in this position that /ŋ/ is not plosive in itself in this variety, if the nasal velar is in the environment of a velar or guttural plosive, the sound-sequences are marked as <nkk>/<nkq>.

To sum up so far, the orthography of the text marks <VqV> grapheme-sequences in positions in which it is not expected. This single orthographic phenomenon, however, converts several phonetic phenomena, which probably end up in results close to each other, a long vowel or diphthong. The tables below summarize the data, and compares markings of the phoneme-sequences. I underlined the phoneme-sequences.

---

<sup>4</sup> Sertkaya (1993) proposes that this word is identical with OT *yanqu* 'echo' (ED 949) with the loss of the word-initial /y-/. The system he draws out in his article about the loss of /y-/ and other phonetic features seem plausible, but in the case of this word the meaning 'echo' is improbable. This word refers to pictures in the text, which are visual depictions of things and not audial. The etymology I propose above, that this word derives from the verbal stem *ay-* with the formative –GU and ultimately means 'reminder' or 'memory' is simpler on one hand, and hasn't got this catachresis on the other, since one of the basic functions of a picture is to remind.

### T.1 Phonemes marked by <VqV> in illabial environment.

Grapheme-sequence	Phoneme-sequence	Lexemes	Instances	Type
<yq̣'>	/yaỵi/	yag̣ir +i	2/4, 2/4	triggered
	/iya/	qoriyan	12/3,14/9, 15/3 (29/7)	word stem
<'q̣'>			17/4 (29/7)	
	/ā/	qa:r	27/9	
		qa:rluy	28/3	
		ta:m	28/6	
		ša:m	33/3	
	/ā/ (?</a/)	qa(:)ṭir	31/3	
<'qy>	/aỵi/	aỵiz +i	(1/6). 13/4,14/5.30/8 (1/6), 13/1	triggered
	/aỵi/	çiray +i	1/5, 34/3	morpheme-boundary

### T.2 Phonemes marked by <VqV> in labial environment.

<'qw>	/aβ/	taβraq	13/6	word stem
	/ay/	tarlay+siz	29/8	
	/γu/	uqγuluγ	35/8	anaptyctic
aγyu		1/1		
<wqw>/<wkw>			6/6	
	/oγu/, /ögü/	oγul (+i/un)	20/3,21/1	triggered
kögüz+ü(+Cx)		1/9, 2/4, 2/5		
<wqy>	/uβi/	suβ+i	19/4	morpheme-boundary

For better transparency, one should try to find one-to-one correspondences between phonemes and graphemes in the above sequences, more precisely, the question is: What phonemes may a grapheme mark in this sequences? T.3 summarises the combinations found in T.1 and T.2

### T.3 Phonemes corresponding to graphemes in <VqV> sequences

< y q ' >			< ' q ' >			< ' q y >			< ' q w >			< w q w > < w k w >			< w q y >		
ya	γ	(i)	i	y	a	a	γ	(i)	a	β	-	-	γ	u	u	β	i
i	y	a	a	-	a	a	y	i	a	γ	-	o	γ	(u)			
			a	γ	(i)				-	γ	u	ö	g	(ü)			

One of the most important features seen in T.1 and T.2 that several words have fluctuating spelling. These are *ayiz*+Px3, *qoriyan* among the illabials, and the formative *-GU* among labials.

May be the most easy to analyse is the <q> element of the sequence. Phonetically it shows a zero element, as it can be seen on the primary long vowels which are marked by <'q'>. In these cases the scribe understood the long vowels as a Mongolian secondary one, the 'hiatus' left by the disappearance of an etymological /g/ is filled by the preceding vowel through secondary lengthening. The cases of <'q'z> *ayiz* 'mouth' seem to confirm this assumption, with the omitting of the second-syllable /i/ and the 'hiatus' left by it causes the preceding vowel to lengthen, thus providing the pronunciation [a:zi]. When the two vowel-marking graphemes mark vowels of different quality, however, we get a different picture. In the cases of <yq'>, we can see diphtongs (or vowels with palatal coarticulation) of different origin. In the case of *qoriyan*, as the Arabic spelling of the word also suggests, we can see an [i̠a] diphtong, in which the [a] element is dominant, and this is the cause why it could also be spelled as <'q'>. In the case of /yay(i)/ the spelling of the word is actually a misspelling, but a systematic one, based on the similar phonetic quality of the sequence <yq'> = [ya:] </yay(i)/ and <yq'> = [i̠a] /iya/.

The <'qy> spelling of *ayiz* shows that the omitted second-syllable vowel leave some trace, thus ending in [a̠]. This ends up in a similar result in the case *čiray+i* where the end of the word already had been a diphtong [a̠], otherwise it should have been spelled as \*<čyr'yy> or, in the case of a stem-final vowel, \*<čyr'sy>.

The overlapping spellings of the different phoneme-sequences suggest that phonetically they are very close to each other, while the instances of individual lexemes, which are spelled differently suggest that the diphtongs containing an [i̠] element are either unstable or their pronunciation are not very far from [a:].

The fact that *čiray+i* has got a syllabic morpheme which is spelled the same way as if it would be a single diphtong rises the question whether these diphtongs should be considered monosyllabic or disyllabic?

In the cases containing <w> we must see a diphtong containing a labial element. The main question here still remains that whether these sequences should be read as mono- or disyllabic? The spellings of *oyul* +Px, *kögüz*+Px and *suβ*+Px may allow a disyllabic reading, but all the other data suggest monosyllabic one, with the labial element being dominant (the 'hiatus' imagined by the scribe filled by the labial element). This would mean that the a last syllable of the word stem, and the possessive suffix, which is syllabic morpheme itself should be counted as a single syllable, which seems to be unlikely.

The pronunciation of <'qw> and <wqw> again must have been very slightly different. This assumption is supported by the following example:



(16) <'wqwz> 1/9 ~ OT *ayuz* 'colostrum, first milk' (ED 98)

The text contains several false etymologies of tribe names (according to the text) originating of given names based on deeds. Although not explicitly, the name of the protagonist also belongs to that type. The name of the protagonist, *Oγuz* <'wqwz>, appears only after the following sentence:

(16a) *ušol oγul ana-si-niŋ kögüzündün ayuz-ni içip mundun song içmädi yig ät aš sorma tilädi*

'That child drank the colostrum from his mother's breast, after that he did not drink anymore. He asked for raw meat, food and wine.'

In this this sentence *ayuz* <'wqwz> 'colostrum', and the spelling of the protagonist's name (which originates from this word) coincides in the text, so the false etymology of the name is based on the coinciding pronunciation of *layul* and *loyul*, thus the case must be similar in the other data cited above as well.

The question still remained that how many syllables a <VqV> or <VkV> sequence to be read, and parallelly another question is to be asked: Is the <VqV> and <VkV> spellings for diphtongs and secondary long vowels relevant only in hypercorrect forms and the data introduced above, or should all <VqV> spellings (for original Turkic /VgV/ sequences as well which do not belong to the *ayiz*-type) to be read similarly?

The data cited above could be identified in the way that the scribe wrote a grapheme-sequence which originally marked two syllables, but in many cases it marks only one in a given word, thus, the number of the syllables of a word is different than it would be expected. The PON shows the form of rhythmic prosa, non-versed and versed parts follow each other, the versed parts can be identified of grammatical or non-grammatical rhyme. The grammatical rhyme is typical feature of the Turkic verse, along with verse-lines consisting of 7, 8, or 11 syllables. In order to solve the above questions, we must identify verses in the PON which contain <VqV> or <VkV> sequences, and count the syllables in the individual lines. Whichever (mono- or disyllabic) reading of the <VqV> and <VkV> sequences fulfill the syllable-number requirement of the versed line, that must be the correct one.

Unfortunately, this method is not without problems. It is not always easy to locate a verse or decide whether it is a verse or not, because the verse-seeming parts do not always fulfill the above requirements. Thus, as far as I have seen, an exhaustive investigation from this point of view is not possible, however, I would like to introduce some examples. The numbers behind the lines show the number of the syllables within the line.

(17) *ay sän munda beglärgä bolyil bašliq (11)*  
*mä mäñiläp sanga at bolswn qayarliy (11)*

' *Oh, here thou shall become leader of the begs!*  
*Lo, be happy and thy name shall be Qärliy!*'

In example (17) the number of the syllables will be correct only if we read the highlighted sequence as one syllable.

- (18) *[ay] [o]ɣwllar köp män (y)ašadum (8)?*  
*urwšqqlar köp män kördwüm (8)*  
*čida basa köp oq attum (8)*  
*ayyir birlä köp yörüdüim (8)*  
*dušmanlarni iylaɣurdwm (8) ?*  
*dostlarumnä män külgürdüim (8)*  
*kök tängriɣä män ötädüm (8) ?*  
*sänlärgä berämän yurtu[m] (8)*

*'Oh sons, I lived much,  
 I saw many battles,  
 I threw many javelins and shot many arrows,  
 I rode much on my stallion.  
 I made my enemies cry,  
 I made my friends laugh.  
 I paid my debt for the Blue Sky,  
 I give my empire for you.'*

In example (18) we see the opposite the causative –GUR- after a stem-final vowel gives the expected number of syllables if we read the /VgV/ sequence as two syllables. The same is true for that in *tängri +gä*. In the first line of this verse, the correct syllable-number will be only if *oɣul+lar* would be read with two syllables. In this case however, it would break the rhythm present in the following lines. In *oɣul+lar* and *iyla-ɣurdum* the /VgV/ sequence stand before a consonant cluster, and it would be possible that the cluster blocks the the monosyllabic pronunciation. In example (19) we see the opposite:

- (19) *qanya qanya birlä ölwgni tirig yörüɣürsün (14)*  
*qanyaluy sanga at bolyluy qanya bälɣürsün (14)*

*Let the living make the lifeless walk with the qanya qanya!  
 Qanyaluy, a name is to be for you, let the qanya manifest (it)!*

In example (19) the number of syllables will be correct only if the highlighted part is read as single syllable, which contradicts the corresponding part in (18).

We can see another instance of *oɣul* in (20):

- (20) *ušol oɣul anasiniŋ (7)*  
*kögüzündün ayuz ni içip (7)*  
*mundwn artiqraq içmädi (8)*  
*yig ät aš soyrma tilädi (7)*  
*tili kelä bašladı (7)*

*qir°q kündün song bädüklädi yöriüdi oynadı (15=8+7)*

*'That child, from his mothers  
breasts, drank the colostrum  
then he did not drink any more,  
he wished for raw meat food and wine,  
he started to speak,  
after forty days, he grew up, walked and played.'*

The rhyme-structure of (20) is *xxaaaa*, not perfect. The syllable number of the lines will be correct only if we consider *all the three* highlighted parts mono-syllabic.

(21) *bir buyu aldı şol buyunı talnung çuβuqı birlä (14=7+7)*

*iyäçqa bayladı ketdi (7)  
andan song ertä boldi (7)  
tang ertä çayda kaldi (7)  
kördi kim <kyynd> buyu nı alwp turwr (9+?)  
kenä bir ađwγ aldı altwnluγ belbayı birlä (14=7+7)  
yiyäçqa bayladı ketdi (7)  
mundwn song ertä boldi (7)  
tang ertä çayda keldi (7)  
kördi kim <d''w> ađwγ nı alip turwr (10+?)*

*'He took a deer. That deer with a willow twig  
He tied to a tree, then went away.  
After that it became tomorrow.  
He came at daybreak,  
and he saw that the [monster] had taken the deer.  
Then he took a bear. With his golden belt,  
He tied it to a tree and went away.  
After this it became tomorrow.  
He came at daybreak,  
and he saw that the [monster] had taken the bear.*

Example (21) is very valuable from the point of view that two paralel episodes follow each other, however, the corresponding lines do not follow exactly the same syntactic sturcture, differ slightly, but the lines still keep the syllable number strictly, keeping forward that the highlighted parts are considered mono-syllabic. It is probable that it is true for the lines in which the '[monster]' occurs, even if the words for '[monster]' has no correct reading so far. Consider however, example (22):

(22) *oγwz qayan baštı (6)  
mıs°r qayan qaçtı (6)  
oγwz anı bastı (6)*

*yurtin aldi ketti (6)*

*'Oγuz Qayan succeeded,  
(the) Qayan (of) Egypt fled.  
Oγuz crushed him,  
he took his empire and went away.'*

(22) is a very strong counterexample for the ones introduced so far. The structure of the lines follows a very strict pattern, and the number of syllables will fit only if every single word are considered disyllabic including *mīs<sup>o</sup>r*, which contains an anaptyctic sound, which is marked in the spelling <m's'r>.

We saw that applied in a verse, that /VgV/ sequences of similar phonetic shapes in similar environments have a dual nature: they can behave as mono- or disyllabic. They can be monosyllabic even if they cross morpheme-boundaries, thus they adapt the syllable number requirement of the verse-line:

monosyllabic		disyllabic
(11) qayarliγ	~	(20) qayan
(17) yörü-gür-	:	(16) iyla-γur
(19) bel+baγ+i	~	(16) tängri+gä
(18) oγuz	:	(20) oγuz

Thus <VqV>, <VkV> sequences mark diphthongs or primary or secondary long vowels on the phonetic level. The marking fluctuates, and there are overlaps between the marking of different diphthong types, allowing the conclusion that their pronunciation is only slightly different if any.

As I mentioned in the introduction, script and language must be considered separately. Script always used by a scribe, and language is used by a speaker. Of course, these two can be the same person during creating a text, but this is not at all necessary. According to the above observations, I assume that in the case of the PON we must assume two different persons as scribe and as speaker.

Hypercorrect forms and fluctuating sound-marking shows us the phonetic level that the scribe *percieves*. These phenomena is due the scribe's inconfidency what he percieves. Thus his phonological basis is different of that of the speaker. The scribe surely had competence in (written and spoken) Mongolian, otherwise he wouldn't have used a Mongolian orthographic feature to mark long vowels. To go one step further, I would risk that his mothertongue was Mongolian, since he percieved primary long vowels as secondary ones developed from two syllables. Thus, his primary phonological basis did not contain only long vowels of this type, while, the Turkic variety he took down seem to preserve those even in this very late stage, even is they had already not been in phonological opposition.

The scribe, by his 'mistakes', provided information of the speaker's phonetic level. The syllabic adaptability of the sounds marked by <VqV> and <VkV> verifies the diphthongic phonetic value, at least on the phonetic level, as [g] and [γ] seem to be not pronounced, the spellings <VqV> and <VkV> otherwise do not containg /g/ in written Mongolian, thus, the

orthography mainly hides this feature. The question still remains whether the speaker's phonological basis still contained intervocalic /g/. His ability to apply /VgV/ as disyllabic sequences supposes that the answer is yes, but on the phonetic level, it must have been close to zero. The speaker's language must have been a Kipchak variety, as the coincidence of the phoneme sequence /aβ/ (5) and /aγ/ (6) supposes.

As a final conclusion, I would correct Pelliot's point with the following: 1. The speaker who told the story was a Kipchak speaker, not the scribe. The scribe knew Written Mongolian, and a spoken variety in which the <VgV> > /VV/ development has already finished, as he spelled long vowels as disyllabic ones. May be this variety was his primary code, but at least he knew a different Turkic dialect than the speaker. 2. The text is not a copy, but a primary one, it had been written after dictation. Pelliot must have been correct to point out the 15th century for the creation of the manuscript, as the Mongolian sound development had already been finished in the variety known by the scribe. The place of the creation of the manuscript may have been the Golden Horde.

## Abbreviations:

ar. – Arabic  
ED – Clauson 1972  
L – Lessing  
OT – Old Turkic  
PON – Pagan Oγuz-nāmā  
Z – Zenker 1866

## References

- Bang, W. 1932. – *Die Legende von Oghuz Qaghan*. Berlin
- Bang, W. 1936. – *Oğuz Kağan Destanı*. İstanbul
- Clauson, Sir G. 1972 – *Etyimological Dictionary of Pre-Thirteenth Century Turkish*. Oxford
- Erdal, M. 1991. – *Old Turkic Word Formation I-II*. Wiesbaden
- Janhunen, J. 2003 – Proto-Mongolic. In: Janhunen, J. (ed.) *The Mongolic Languages*. London, New York. 1-27.
- Lessing, F. D. 1973. – *Mongolian-English Dictionary*. Bloomington
- Miller, R. A. 2002 – The Middle-Mongolian vocalic hiatus. In: *Acta Orientalia Hungarica* 55. 179-205
- MS No 1001*. Bibliothèque Nationale, Supplément Turc. Paris
- Nour, R. 1928. – *Oughuz-namé, épopée turque*. Alexandrie
- Pelliot, P. 1930. – Sur la légende d'Uγuz-Khan en écriture ouigoure. In: *T'oung Pao* XXVII. 247-358.
- Pelliot, P. 1995 – *Uygur yazısıyla yazılmış Uğuz Destanı üzerine*. (çeviren: Vedat Köken) Ankara
- Radloff, W. 1891 – *Das Kudatku Bilik von Jusuf Chass-Hadschib aus Balasagun*. St. Petersburg. 232-244.
- Ščerbak, A. M. 1959. – *Oguz-nāme; Muḡabbat-nāme*. Moskva
- Sertkaya, O. F. 1993 – Einige neue Lesungen und Ineterpretationsvorschläge zur Legende von Oghuz-Kaghan. In: *Altorientalishce Forschungen* 20/2 360-368
- Zenker, J. T. 1866 – *Dictionnaire Turc-Arabe-Persan*. Leipzig