Offprint

RELIGIOUS AND LAY SYMBOLISM IN THE ALTAIC WORLD AND OTHER PAPERS

Proceedings of the
27th Meeting of the Permanent International Altaistic Conference,
Walberberg, Federal Republic of Germany
June 12th to 17th, 1984

Edited by
Klaus Sagaster
in collaboration with
Helmut Eimer



1989
OTTO HARRASSOWITZ · WIESBADEN

Doug Hitch (Dundas, Canada)

OLD TURKIC ‚*b*' AND ‚*p*'[1]

0. Introduction

A difficult area of Old Turkic phonology has concerned the labial consonants. seems to be most generally agreed that the language had three such phonem /*b*, *p*, *v*/.[2] The Brāhmī and Manichaean scripts have enough graphs to represe these sounds and it is thought that in the Uygur/Sogdian script *p* represents bo stops /*b*, *p*/ while in the Runic b^1, b^2 represent both voiced segments /*b*, *v* There are at least two divergent opinions in the literature. Omeljan Pritsak fee

1 Abbreviations.
ATG: A. v. Gabain, *Alttürkische Grammatik* (Porta Linguarum Orientalium XXIII), 3rd ed., Leipzig 1974.
Clauson, *Origin*: Sir Gerard Clauson, *The Origin of the Turkish „Runic" Alphabet*, Acta Orientalia 32, 1970 (XI PIAC, Hørsholm Denmark, June 2-6, 1968), p. 51-76.
EDPT: Sir Gerard Clauson, *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*, Oxford 1972.
Erdal, *Chron. Class.*: Marcel Erdal, *The Chronological Classification of Old Turkish Texts*, CAJ 23, 1979, p. 151-175.
Hybrid: Doug Hitch, *A Brāhmī-Manichaean Hybrid Script*, Orientalia Lovaniensia Periodica 14, 1983, p. 293-312.
v. LeCoq, *Köktürkisches*: A. v. LeCoq, *Köktürkisches aus Turfan*, SKPAW 1909, Phil.-Hist. Kl., p. 1047-1061 and plates IX, X, XI and XII.
Maue, *Bilinguen*: Dieter Maue, *Sanskrit-Uigurische Bilinguen aus den Berliner Turfanfunden*, Giessen 1983.
Orkhon Turkic: Talât Tekin, *A Grammar of Orkhon Turkic* (UAS 69), Bloomington/The Hague 1968.
Pritsak, *Alttürkische*: Omeljan Pritsak, *Das Alttürkische*, in: Turkologie (Handbuch der Orientalistik, 1ste Abt., 5de Bd., 1ster Abschnitt), Leiden/Köln 1963, p. 27-52.
Röhrborn, *UW*: Klaus Röhrborn, *Uigurisches Wörterbuch*, Lieferung 1, Wiesbaden 1977.
TMS: Sir Gerard Clauson, *Turkish and Mongolian studies*, London 1962.
TTT II: W. Bang and A. v. Gabain, *Türkische Turfan-Texte. II.* SPAW 1929, Phil.-Hist. Kl., p. 411-430 and plates V and VI.
TTT III: W. Bang and A. v. Gabain, *Türkische Turfan-Texte. III*, SPAW 1930, Phil.-Hist. Kl., p. 183-211 and plates III and IV.
TTT VIII: A. v. Gabain, *Türkische Turfan-Texte VIII*, ADAW Kl. Spr. Lit. Kunst 1954, 106 pages and 2 plates.

2 Not all descriptions are phonological. A recent state of the art exposition is found in Röhrborn, *UW* p. 13-17.

that *v* stands for an allophone of /*b*/[3] while for Talât Tekin, Orkhon Turkic features the ‚Preservation of the bilabial stop *b* which later developed into *w* [i.e. *v*] in medial and final positions'.[4]

1. The problem

I have recently been conducting research into the Brāhmī scripts once used for writing a variety of languages spoken along the silk route.[5] It is clear that the Old Turkic Brāhmī materials are more coherent and useful than is widely assumed. At one point, with the purpose of ascertaining the role of *pa*, I surveyed the labial stop spellings in the documents. It was rather surprising to find that although five labial plosive radicals were in use, *pa*, *pha*, *ba*, *bha* and the new sign *pa*, these are distributed in a fashion that would preclude rather than justify the positing of two contrasting stop phonemes */*b*~*p*/.[6] In the majority of the documents *pa* appears in virāma, an orthographic position, while *pa* occurs in nonvirāma (± elsewhere). The implication of this is that the language had a single labial plosive phoneme. This observation contravenes traditional Old Turkic thinking and as the point is essential for other studies, it was necessary to pursue it as far as possible.

Pritsak's hypothesis seems to be that the language had two labial phonemes, voiceless /*p*/ and voiced /*b*/, the latter having an initial allophone [*b*] and a noninitial realization [*v*].[7] This view can accommodate the evidence of both the Manichaean and the Runic scripts. In Manichaean writing *b* appears only initially while *p* and *v* appear only noninitially.[8] Pritsak would argue that *b* and *v* represent the allophones of /*b*/. In Runic writing b^1, b^2 occur in all positions while *p* is restricted to noninitial position. For Pritsak these spellings are completely phonemic. However, the new Brāhmī evidence cannot be accommodated to this hypothesis. It allows the positing of only one labial plosive written predominantly *pa*/*pa* and which contrasts with the fricative written *wa*.

3 Pritsak, *Alttürkische*, p. 32, par. 9.1, n. 2. See also here below, fn. 7.
4 *Orkhon Turkic*, p. 7.
5 (1) *Hybrid*. (2) *Central Asian Brahmi Palaeography: The Relationships among the Tocharian, Khotanese and Old Turkic Gupta Scripts*, unpublished MA thesis, University of Calgary 1981. (3) *Kharoṣṭhī Influences on the Saka Brāhmī Scripts*, in: W. Skalmowski and A v. Tongerloo, eds., Middle Iranian Studies (Orientalia Lovaniensia Analecta 16), Leuven 1984, p. 187-202.
6 See below, *13. Brāhmī script*.
7 The actual statement is: ‚Auch das „uig." *-w-*, *-w* und das Brāhmī *-ð-* sind Allophone, und zwar von /*B*/ und /*D*/." See above, fn. 3.
8 See below, *12. Manichaean script*.

2. *Theory*

One way to accommodate the conflicting kinds of evidence is to posit a phonological change within Old Turkic. That is, originally the language had /*b*/ and /*p*/ but no */*v*/, and later noninitial /*b*/ fricativized to [*v*]. So far this harmonizes with T. Tekin's view. One needs then to go a step further and posit the merger of original initial /*b*/ with original noninitial /*p*/ and the simultaneous erection of the new contrasting unit /*v*/. That is, after the shift the language had one plosive and one fricative phoneme replacing the earlier two stops. This formulation has in common with the Pritsak hypothesis that Old Turkic had just two labial consonants, but allows instead for the existence of two disparate phonological structures.

The theory has numerous advantages. One can now clearly and easily explain the various orthographic categories involving labials in all the Old Turkic scripts. Two phonological phenomena affecting the labials, *b*>*m* regressive assimilation and *v*/*p* alternations in clusters can now be motivated. A new isogloss for Old Turkic dialect studies becomes available. Finally, this provides a new tool for research into labials problems in mixed materials and in other Inner Asian languages and scripts – but that research is not the purpose here.

3. *Dialectology and nomenclature*

Old Turkic dialectology and the nomenclature appropriate for the dialects are topics beyond my abilities. Nevertheless, some labels are required for the purposes of exposition. After much struggle and with great reservations I use here the strictly relative terms of Earlier Old Turkic (EOT) and Later Old Turkic (LOT) to refer to the stages respectively before and after a particular phonemic change. Several such innovations are discussed in these pages, not only the labials shift, but also the better known processes of *ñ*>*y* or *n*, and *b*>*m*. At a very early stage none of the changes had taken place and at a very late stage all had occurred. But in between these stages it is not yet determined how the processes chronologically or geographically relate to each other. EOT and LOT do not refer to well-defined, uniform dialects but to grammatical structures before and after the particular change under discussion.

4. *Sketching the shift*

The theory of a consonant shift is partly sketched in Chart One. These are the changes which affect the majority of the segments. Special considerations are to be made later for words showing *b*>*m* assimilation[9] or *p*/*v* alternations in clusters.[10] The chart works as follows: At the top of each column are phonetic values in square brackets organized according to linguistic position. Within each column are phonemes having the phonetic realization listed at the top. Three stages are listed, each with a different phone~phoneme relationship. The uppermost stage shows original phonetic and phonemic values. The phonetic shift shows a change in phonetic values and the bottom stage illustrates the phonemic reorganization of the new phonetic material.

In EOT we find /*b*/ pronounced [*b*] and /*p*/ pronounced [*p*]. Then a phonetic shift takes place which affects the noninitial allophones. Original /*b*/ fricativizes to [*v*] and /*p*/ voices to [*b*]. (These parallel and more or less simultaneous changes are both manifestations of a more general pattern of consonant weakening in Old Turkic.[11]) After the phonetic shifting is complete, the now phonetically identical reflexes of EOT initial /*b*/ and noninitial /*p*/ merge as LOT /*b*/. At the same time, original noninitial /*b*/ has become the new phoneme /*v*/ in LOT.

Chart One: The Old Turkic labial consonant shift

	noninitial position			initial position
(phonetic values =)	[v]	[b]	[p]	[b]
EOT	–	/b/	/p/	/b/
	↙		↙	
phonetic shift	/b/	/p/	–	/b/
phonemic reanalysis, LOT	/v/	/b/	–	/b/

It will now be argued that the attested Old Turkic materials conform to this pattern as well as provide evidence for it. This is done more or less according to the three stages in the chart, beginning with the evidence that EOT had /*b*/ and /*p*/ but no */*v*/.

9 See below, *9. Initial m.*
10 See below, *11. Clusters containing labials.*
11 See below, *8. Old Turkic consonant weakening.*

5. *Structure of the Runic script*

The breakdown of the orthographic categories recognized in the Runic script suggests that the dialect to which the script was first adapted[12] did not possess a /v/. All consonant phonemes have unique representation in this system.[13] GERARD CLAUSON was of the opinion that several graphs were ‚polyphonic'.[14] He did not recognize the phonemic principle and attributed ‚polyphonicity' to those four graphs which stood for more than one sound: ‚*p*' stood for ‚*p*/*f*', ‚*d*' for ‚*d*/*ð*', ‚*q*' for ‚*k*/*x*' and ‚*b*' for ‚*b*/*v*'. Today Old Turkic ⟨*f*⟩[15] and ⟨*x*⟩ are regarded as foreign phonemes while Old Turkic [*ð*] is analysed as an allophone of /*d*/.[16] Just ‚*b*' is still thought to represent more than one phoneme.

The fact that no */*v*/ category exists in Runic writing is not an oversight. I agree entirely with CLAUSON that the inventor ‚had high intellectual and educational qualifications'.[17] His theory that the script is a Graeco-Iranian hybrid is in many essentials a plausible one.[18] If it is correct, then we should be impressed that letters for /*e*, *ñ*, *ŋ*/ were not borrowed but invented.[19] Even the Uygur/Sogdian and Manichaean scripts do not have special graphs for these phonemes. The inventor further created *γ* and *š*[2] as the back counterparts of *g* and *š*[1]. All this points to a high degree of linguistic sophistication.[20] If a */*v*/ existed originally in Old Turkic a Runic graph or pair of graphs would have been borrowed or invented in order to write it.

12 This dialect is not necessarily the same as Orkhon Turkic.

13 It might be objected that the sibilants are not uniquely represented. But while it is true that no single documented system uses four basic sibilant letters (some paper materials, however, derive two palatal signs from the two dentals by means of a diacritic: v. LECOQ, *Köktürkisches*, p. 1054), four separate and morphologically distinct sibilant graphs are attested in the materials taken as a whole (CLAUSON, *Origin*, p. 64-5 and 74-5). In my opinion a proto-system of sibilant spelling can be reconstructed from which all the transmitted systems can be logically derived.

14 *TMS*, p. 77-8.

15 Here I introduce the practice of putting foreign phonemes in angled brackets. It is useful to give a special transcription to foreign and marginal phonemes since these often play a fundamental role in historical change and in the structure of scripts. The usefulness is to be seen below in *9.0 Initial m*, where the relevant distinctions cannot be diagrammed without a special device like this.

16 For instance see RÖHRBORN, *UW*, p. 10-11, and PRITSAK, *Alttürkische*, p. 32.

17 *TMS*, p. 73.

18 CLAUSON, *Origin*.

19 *Ibid.*, p. 68-70.

20 *Ibid.*, p. 69-70.

6. *Runic dialects*

It seems to be a widely accepted principle that major dialectal differences exist or could exist within the attested Old Turkic materials in the Runic script.[21] At the same time it is agreed that an important isogloss within the language is still the one first noted by A. v. GABAIN: older *ñ* versus younger *y* or *n*.[22] Any Runic text containing the graph *ñ* is automatically ascribed to the *ñ*-dialect.

In the famous article *Köktürkisches aus Turfan*[23], ALBERT v. LECOQ published some valuable mixed materials. These are Runic script texts with Manichaean content and containing either a Western Middle Iranian language (Parthian or Middle Persian) or both Iranian and Old Turkic. A unique document is an alphabet chart in which Runic letters are transcribed in the Manichaean script.[24] Number 10 on this chart, , is transcribed in Manichaean script as *iy*.[25] At that early stage of Old Turkic studies (1909) it was not yet known that a palatal nasal consonant once existed and so v. LECOQ followed V. THOMSEN's practice and transcribed the Runic letter as *i̯* without further comment. As far as I can tell, no one has noticed that this letter is in fact to be identified as *ñ*; that is, Runic *ñ* = Manichaean *iy*. This entails the conclusion that the Turk who wrote this chart spoke a dialect in which /*ñ*/ had already become /*y*/. Nevertheless, he still knew about the existence of the graph *ñ*. Another manuscript from the same publication in Runic and Old Turkic with Iranian rubrics, T. M. 342, was found only about 15 kilometers away from the alphabet chart.[26] It is reasonable to assume that both pieces are a product of a single linguistic and orthographic tradition. On this fragment the word *ayıγ*/*anıγ*/*añıγ* ‚bad, evil' occurs with *ñ* used ‚correctly'. In the whole corpus of the runic materials the graph *ñ* appears in only 8 morphemes[27] and it would not have been difficult for a scribe who spoke a *y*-dialect to learn to write *ñ* ‚correctly' some-

21 For instance see PRITSAK, *Alttürkische*, p. 27-28, *TMS*, p. 68-72. ERDAL, *Chron. Class.*, has most Runic texts in group I, but at least one is included in group II, p. 165.

22 *ATG*, p. 3-8. Followed for instance by: *TMS*, p. 117-118; PRITSAK, *Alttürkische*, p. 29; ERDAL, *Chron. Class.*, p. 152.

23 v. LECOQ, *Köktürkisches*. Apparently reprinted in: *Sprachwissenschaftliche Ergebnisse der deutschen Turfan-Forschung*, Bd. I, Leipzig 1972, p. 532ff. These texts have also apparently been re-edited in a work to which I have not had access: volume 2 of H. N. ORKUN, *Eski Türk Yazıtları*, 3 volumes and index, Istanbul 1936-41. I have also not been able to see the re-edition of the Parthian texts in: C. SALEMANN, *Manichaica III*, Bull. de l'Ac. Imp. des Sc. de St. Pétersbourg, 1912, p. 1-32.

24 v. LECOQ, *Köktürkisches*, p. 1050, signature: T. II, T. 20.

25 That is, *aleph-yod-yod*. According to v. LECOQ this was corrected from *aleph-aleph-yod*, ibid.

26 *Ibid.*, p. 1048.

27 *TMS*, p. 91.

times as part of his orthographic training. It must be concluded that the presence of the letter $\tilde{n}$ in a runic text does not necessarily mean that the author had a */ñ/ phoneme in his speech. Neither is it a safe indication of the relative age of the language in a manuscript.

7. *Foreign* ⟨*f*⟩

I have noticed that the foreign phoneme ⟨*f*⟩ is treated in two different ways in two sets of Turkic Runic materials. In the Orkhon Turkic inscriptions the sound is twice written with *p*. T. Tekin transcribes the words in question as *purum* and *apar*[28] while Clauson transcribes *Forom* and *Afar*.[29] According to Tekin the first word comes from Parthian *frum* ‚Rome' while the second is thought to be the national name ‚Avar'. In contrast to this Orkhon Turkic system, the paper fragments from v. LeCoq's article[30] feature Western Middle Iranian /*f*/ always written with b^1, b^2. The four Runic spellings in question are here compared with the phonemic transcriptions given in Mary Boyce's Wordlist:[31]

$b^2(i)r^2iš^2[t^2]///$ (T. M. 327, r. 11) Parthian */frēštag/* ‚apostle, angel'
$ab^1r^1in^1$- (T. M. 327. v. 6) Parthian */afrin-/* ‚invoke blessing, pray'
ab^1ur^1- (T. M. 339a. v. 1) Middle Persian */āfur-/* ‚invoke blessing, bless, pray'
$i(?)b^1r^1as^1$ (T. M. 342. v. heading) Parthian */wifrās/* ‚teaching, instruction'

The disparate ways of writing ⟨*f*⟩ in the runic materials, with *p* in Orkhon Turkic and with b^1, b^2 in the paper documents, result from a difference in dialect. In the EOT of Orkhon Turkic the phoneme phonetically closest to ⟨*f*⟩ was /*p*/ since a */*v*/ did not exist. Similarly, in the LOT of v. LeCoq's materials the phoneme phonetically closest to ⟨*f*⟩ was /*v*/ and a */*p*/ did not exist. In both sets of documents the scribe has simply chosen to write the foreign sound with the Runic graph having the closest phonetic value. In this way we know that the mere presence of the graph *p* in a runic document, does not guarantee that a */*p*/ existed, just as $\tilde{n}$ does not prove */ñ/.

Other facts from the paper materials support the conclusion that b^1, b^2 = /*v*/. The Turks using the script show a high degree of linguistic sophistication. They employed a diacritic bar or dot to modify the values of four runic graphs.[32] Two

28 *Orkhon Turkic*, p. 75.
29 *TMS*, p. 77.
30 See above, fn. 23.
31 Mary Boyce, *A Wordlist of Manichaean Middle Persian and Parthian*, (Acta Iranica III. II. Supplement) Leiden 1977.
32 v. LeCoq, *Köktürkisches*, p. 1054-5, 1059.

of these modified letters were used for the foreign Iranian sounds: $\bar{\gamma}$=⟨*h*⟩ and $\bar{z}$=⟨$\dot{j}$⟩[33], cf. $y^1ožt^1\bar{\gamma}///$ (T. M. 339b). Middle Persian */yōj̇dahr/* ‚holy, hallowed'.[34] The other two letters were developed to improve the writing of the Turkic sibilants. The diacritic was drawn above s^1 and s^2 to give the corresponding palatals: $\bar{s}^1=š^1$, $\bar{s}^2=š^2$.[35] In the majority of the Runic materials four sibilants are not distinguished.[36] The scribes here showed a high degree of sensitivity towards the finer points of phonological and orthographic detail. If the Turkic dialect had been EOT one would expect to find an innovation like *$\bar{p}$=⟨*f*⟩. The fact that b^1, b^2 were adequate for writing foreign ⟨*f*⟩ must mean that the graphs stood for the fricative /*v*/.

8. *Old Turkic consonant weakening*

Consonant shifts of the kind EOT *p* > LOT *b*, and EOT noninitial *b* > LOT *v*, fit well within the general pattern of sound change in Old Turkic. There is evidence from spelling interchanges that under certain conditions voiceless consonants voiced, and that voiced stops became fricatives. Because these developments progress with time, Marcel Erdal has used them as a means of chronologically classifying the Old Turkic texts.[37] In his group I the consonants are rendered consistently. In II there is rare interchange between /*d*/ and /*t*/ and between /*z*/ and /*s*/. Group III shows frequent *d*~*t* and *s*~*z* interchange in all alphabets and *k*~*g* interchange in the Brāhmī and Manichaean scripts. In group IV there is occasional passage of /*d*/ to /*y*/. This last development presumably entailed an intermediate phonetic fricative stage [*ð*].

Although there is nothing phonetically unusual about the labial shifts proposed here, they led to a phonemic reanalysis which is unique. A similar phonological development did not take place with the dentals or gutturals. The reason for the disparity in treatment has to do with the distribution of these consonants. In Old Turkic there are four consonant pairs which are distinguished by a correlation of voicing: /*b*~*p*, *d*~*t*, *z*~*s*, *g*~*k*/. Only one member of each pair appears in initial position but this phenomenon is not governed by the voicing feature. /*b*, *t*, *s*, *k*/ occur initially while /*p*, *d*, *z*, *g*/ do not. When a noninitial voiceless consonant other than /*p*/ tends to voice, it becomes phonetically

33 The Iranian sound may have been a fricative ⟨$\dot{z}$⟩ rather than an affricate ⟨$\dot{j}$⟩. Mary Boyce, *A Reader in Manichaean Middle Persian and Parthian*, (Acta Iranica III. II) Leiden 1975, p. 16.
34 See above, fn. 31.
35 v. LeCoq, *Köktürkisches*, p. 1054.
36 Clauson, *Origin*, p. 64-5. See also above, fn. 13.
37 Erdal, *Chron. Class.*

less like the corresponding initial. That is, for instance, when noninitial /t/ > [d] the phonetic make-up moves away from the [t] of initial /t/. But on the other hand, when noninitial /p/ > [b], it *approximates* the phonetic quality of the corresponding initial, /b/. In this way, the disparate initial distribution among the consonant pairs was conducive to phonemic change among the labials but not among the other categories.

It is convenient to mention here that the phonemic restructuring is also responsible for the fact that we have *t*~*d*, *s*~*z* and *k*~*g* but not **p*~*b* spelling alternations. As is well known, phonologically oriented writing systems always operate primarily from the phonemic level and only secondarily if at all do they recognize further phonetic detail. Because LOT had two dental and two guttural stop phonemes as well as two sibilant phonemes, we find alternations written in those scripts which have graphs for both the voiced and voiceless segments. In contrast, because LOT had just a single labial stop phoneme, no significant alternations occur even when as many as five graphs were available as in the Brāhmī.

9. *Initial m*

Originally in Old Turkic there was no initial /m/. A labial nasal sound did occur at the beginning of foreign words[38] but this must be regarded as a foreign phoneme, i.e. initial ⟨m⟩. At some point, as is well known, initial /b/ became /m/ in words containing nasals. Within the Orkhon Turkic materials the nasalization is beginning. It is not very common, being attested only in the alternations *män*~*bän* ‚I', *maŋa*~*baŋa* ‚to me' and *mäniŋ*~*bäniŋ* ‚my'.[39] The same modification probably happened medially as well. The corpus is just too small for there to be many words with a medial *b* and a following nasal. Such words have to be at least two syllables long. Nevertheless, in Orkhon Turkic the alternation is found once medially in *tarban-qa*~*tarman-qa* ‚geographical name'.[40]

The shift /b/ > /m/ is parallel and related to the /b/ > /v/ shift under discussion here. The two historical developments are compared in Chart Two. The chart has four columns defined by whether or not a nasal follows in the same word, and by linguistic position. Probably at around the same time as the phonetic shift of /b/ > [v] (column 2), in words with a following nasal there was a phonetic shift of /b/ > [m] (columns 3 and 4). In initial position this meant merger of

38 See *TMS*, p. 171 and *Orkhon Turkic*, p. 76.
39 *Orkhon Turkic*, p. 92.
40 *Orkhon Turkic*, p. 93.

the foreign phoneme ⟨m⟩ with the new [m] allophone of /b/ (column 3, phonetic shift stage). Medially, the nasalized /b/ merged with the native phoneme /m/ (column 4, phonetic shift stage). The fact that an [m] sound now existed initially in native words caused all initial [m] sounds, whether originally native or foreign, to be reanalysed as members of a native initial /m/ (columns 1 and 3, phonemic reanalysis).

Chart Two: Old Turkic /b/, ⟨m⟩ > /m/

	not followed by nasal				followed by nasal			
	initial		noninitial		initial		medial	
	1		2		3		4	
(phonetic values =)	[b]	[m]	[b]	[v]	[b]	[m]	[b]	[m]
EOT	/b/	⟨m⟩	/b/	–	/b/	⟨m⟩	/b/	/m/
			↘			↘		↘
phonetic shift	/b/	⟨m⟩	(/p/)	/b/	–	/b/	(/p/)	/m/
phonemic reanalysis, LOT	/b/	/m/	(/b/)	/v/	–	/m/	(/b/)	/m/

10. *Perso-Arabic script*

The structure of the Perso-Arabic script not only shows that LOT had just one labial stop phoneme, it also demonstrates that the result of the merger of EOT initial /b/ with EOT noninitial /p/ was in LOT a phonemically voiced rather than voiceless segment.

Clauson was certainly correct when he wrote: ‚Kāşğarī was a competent philologist ... [who] made every effort to write Turkish words as exactly as possible.'[41] The eleventh century scholar had a highly refined system of writing native as well as foreign consonants. By means of the triple dot diacritic he distinguished native /v/ (*fā* triply dotted) from foreign ⟨f⟩ (plain *fā*), and native /z/ (plain *zayn*) from foreign ⟨ž⟩ (triply dotted *zayn*). His script possessed separate letters for foreign ⟨h, w, x⟩ and for the [ð] allophone of /d/. His perception and description of sounds is also impressive. Kashgari heard and commented on the distinction of voicing among consonants. It is not surprising that he knew about the existence of a voiceless [p] sound, which he called *hard bā* and which he mentioned very occasionally in reference to particular words.[42] Given this degree of phonological competence it should be considered significant that

41 *TMS*, p. 119.
42 *TMS*, p. 121.

Kashgari employed only the voiced letter *bā* to write labial stops and never the voiceless Persian letter triply dotted *bā*.[43] Clearly he regarded his single labial stop phoneme as voiced.

11. Clusters containing labials

Something phonologically peculiar about Old Turkic clusters containing labials has been recognized for a long time. For instance, the first edition of the ATG has a paragraph: ‚Wechsel inschr. *-b-*, *-b* / handschr. *w*; vor Konsonanten auch *p*.'[44] There v. Gabain listed the Runic spellings *qabšur-* ‚collect' and *köbrüg* ‚bridge' together with *-v-* and *-p-* alternates from other scripts. More recently Georg Kara and Peter Zieme make the interesting comment: ‚Wechselnde Schreibungen von *-b-* ~ *-v-* lassen einen phonetisch bedingten Hintergrund vermuten.'[45] They contribute further the alternations *yalbar-* ~ *yalvar-* ‚implore, beseech' and *pinbat* ~ *pinvat* ‚alms'.

As far as I can tell, no one has yet attempted to formulate or to explain the range of spellings. Within the context of the theory presented here, an explanation and a formulation present themselves. The phonetic weakening of EOT noninitial /*b*/ would have been affected by the phonetic environment. In general, intervocalic consonants are more susceptible to weakening than are those in clusters. But it is not simply the case that EOT /*b*/ remained a stop in all clusters everywhere. Evidently, as the alternations suggest, sometimes the original stop remained, and sometimes it fricativized. The factor determining which course was taken could not have been phonetic. The only explanation can be that there were two different LOT dialects, one permitting fricativization in certain clusters and the other not. This idea matches the notion of the development of EOT /*ñ*/ on the one hand to /*n*/ and on the other to /*y*/ in the so-called *n-* and *y*-dialects of LOT. If such an explanation is correct, then a second LOT dialectal isogloss is available for use.

An exact formulation of the relevant dialectal or phonological conditions is not the main point of the present discussion and would most certainly be more effectively accomplished by a professional Turkologist. Nevertheless, with the present theory in hand it is a simple task to recognize other cases of alternations of this kind and to construct a tentative formula. For instance, a brief and UNsystematic search in the *EDPT* produced 5 further examples, not all recognized by Clauson as alternations:

43 *Ibid.*

44 *ATG*, 1941 edition, par. 29, p. 53.

45 Georg Kara–Peter Zieme, *Fragmente tantrischer Werke in uigurischer Übersetzung (= Berliner Turfantexte VII)*, Berlin 1976, p. 8.

öpkä/övkä ‚lung, anger'.[46]
kavsa ‚to surround' „By XI it had become corrupted to *kapsa*:-"[47]
yapšur- ‚to fasten, stick.' Clauson transcribes this once with *-f-* (Man. TT. II 127-8) and mentions that Kashgari specified that *al-fā* was possible.[48]
yelvi ‚sorcery, witchcraft.'[49] Clauson surmises a semantic relationship with the verb *yelpi-*, not recorded before the ‚medieval period' but he commented that „it is difficult to suggest a morphological connection."[50]
yelpik[51] and *yelvik*[52] are not related by Clauson but must be under the present theory. If Clauson's meaning of the verb, ‚to practise sorcery', is correct, then the noun does not mean ‚evil spirits' as he suggests but rather ‚sorceror, witch'.

As can be seen, the theory not only explains the phonological alternations but provides new morphological and semantic insights.

Based on the few examples listed in this paper it is now possible to provide a tentative formula which defines the distribution of the alternations: EOT noninitial /*b*/ in certain LOT dialects remained /*b*/ if it occurred either before /*k, r, s, š*/ or after /*l, n*/.

12. Manichaean script

The Manichaean Turks inherited from the Sogdians two different letters for labial stops. Rather than use only one graph for their single phoneme they developed an orthographic rule to determine when a letter should be used. *Beth* was written at the beginning of a word, and *pe* elsewhere.[53] There are a few instances of noninitial undotted *beth* in the documents but these can probably mostly be explained as loanwords or as cases of the scribe forgetting the diacritic dots which converted the ‚*b*' to a ‚*v*'.[54] It is further to be expected that a scribe who does not phonemically distinguish between two letters will confuse them

46 *EDPT*, p. 9a.
47 *EDPT*, p. 589a.
48 *EDPT*, p. 881a-b.
49 *EDPT*, p. 919b.
50 *EDPT*, p. 920a.
51 *EDPT*, p. 920a.
52 *EDPT*, p. 920b.
53 It is plausible that Manichaean Turks were familiar with the correct employment of Runic writing before developing a version of the Manichaean script for the rendering of LOT. Because initial and noninitial /*b*/ are written with Runic b^1, b^2 and *p* respectively, they came to be written with Manichaean *beth* and *pe* respectively.
54 See for instance *TTT II*, p. 423, where a case of the diacritic being used incorrectly over initial *beth* is also mentioned.

sometimes. There is little doubt that the attested use of *pe* sporadically in initial position[55] is the result of this confusion. There must also have been ‚incorrect' employment of *beth* for LOT noninitial /*b*/.[56]

13. Brāhmī script

Is seems to have been the general opinion that the Brāhmī used to write Old Turkic is chaotic and inconsistent. Two methodological errors lie at the root of this opinion. First, the possibility has not been admitted that different orthographic systems may have been used in different manuscripts. Second, it has not been realized that orthographic as well as linguistic rules are sometimes required to explain spelling patterns. I have recently pointed out that variations like *a͡ya*~*e͡ya* for /*ä*/ or *o͡ya*~*o͡yo* for /*ö*/ in the construction of vocalic radicals are manuscript dependent.[57] In the same place I showed that recognition of the orthographic principle of virāma specialization is necessary to explain the distribution of the guttural signs.[58]

I have conducted a study of the labial stop spellings in the texts in *TTT VIII* and in the further fragments found in Dieter Maue's recent work on the previously unedited Old Turkic – Sanskrit bilinguals in Brāhmī script.[59] The sample initially consisted of all the words containing one or more of the five radicals *pa, pha, ba, bha, pa*. From *TTT VIII* these were collected from the Glossary and from the list of cases of the gerund suffix -*ᵒp*.[60] From Maue's work the words were collected from the transliterations. From this material all foreign or potentially foreign words were removed. Then the instances of the graphs were calculated according to manuscript, to linguistic position and to orthographic position. Theoretically this could entail 810 categories: 5 graphs × 27 manuscripts × 3 linguistic positions × 2 orthographic positions. But because the orthographic distinction occurs in this case only in linguistically final position, because not all graphs occur in each position, and because some manuscripts do not have any of these graphs, only 225 categories need reporting: 25 manuscripts × (4 initial spellings + 3 medial spellings + 2 final spellings). These are diagrammed in Chart Three.

55 See for instance *TTT III*, p. 184, where it is also stated that *f*, otherwise reserved for ⟨*f*⟩, is used sporadically for /*v*/. This is another example of scribes confusing a nonphonemic distinction.

56 I have been unable to find any certain cases of noninitial *beth* where a plosive was intended. The corpus is small and no examples may ever come to light.

57 *Hybrid*, p. 303-4.

58 *Hybrid*, p. 300-302.

59 Maue, *Bilinguen*.

60 *TTT VIII*, p. 80-81.

Since it is generally agreed that Old Turkic had only initial /*b*/ but no initial */*p*/[61], it follows that whatever graphs are used in the Brāhmī for initial labial plosives, the underlying phoneme is always /*b*/. Therefore, in order to look for linguistically significant patterns in these statistics we must first determine the systems used to spell initial /*b*/, and then we can compare the medial and final spellings with these systems.

13.1 Initial spellings

To write initial /*b*/ clearly three and possibly four systems were in use. The most common method featured *pa* as the regular or nearly regular spelling (A, B, C, D, E, G, N, O, P, 1, 3?, 4?, 7, 9?).[62] The next most frequent system has *bha* as the favoured radical (F?, H, K, 6, 10?, 12?) but this pattern is less regular than the preceding one. A third formula exhibits *ba* and *pa* almost equally (L, M, 8?, 11?) and in one document we find *bha* in around three-quarters and *pa* in one quarter of the cases (I). There are no labial radicals in two fragments (2, 5).

13.2 Medial spellings

On the whole, the medial spellings accord with the systems exhibited per manuscript for writing initial /*b*/. In the vast majority of the cases the labial stop is written with *pa* and in almost all these documents, *pa* is also a regular initial spelling (A, G, I, L, M, O, 1, 11). Unfortunately, there are no medial spellings in those manuscripts in which *bha* was the favourite initial graph. The situation in three documents deserves comment. In C *qapıγ* ‚gate' occurs twice with *pha* which might be interpreted as denoting a contrast with initial *pa*. But among the many irregularities in these materials we do find *pha* sometimes as an initial radical (A, K, L). Besides this, *qapıγ* is also found written with *pa* in document L where that graph is in fact a regular initial one. F contains only two initials, both *bha*, against which contrasts the medial *pa* in *yipkä* ‚little thread'. The document is too small perhaps for one to identify an initial spelling system. O has regularly initial *pa*, medial *pa* in *yılpıγ* ‚sprite' and medial *bha* in *qoburγa* ‚owl'. The first medial matches the initial system, and while the second does not, it is unlikely that *bha* would be employed to show voicelessness.

61 A contrasting opinion in a framework that does not recognize the phonemic principle is found in Sir Gerard Clauson's *The Initial Labial Sounds in the Turkish Languages*, BSOAS 24, 1961, p. 298-306.

62 Obviously, the shorter the manuscript, the less valid the observation. Question marks (?) show the fragments which are possibly too short to show significant patterns.

Chart Three: Brāhmī labial stop spellings

TTT VIII

MAUE, *Bilinguen*

Initial Position

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	1	3	4	6	7	8	9	10	11	12
pa	75	12	28	13	23		18		6	1	38	10	11	14	5	9	1	1	3	4	3	2	1	2	1
pha	1									3	1														
ba				3				4		2	31	6								1	4			2	1
bha						2		22	17	13					1				18	1	2	1	2	1	2

Medial Position

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	1	3	4	6	7	8	9	10	11	12
pa	1					1	2		2		1	2		1		3									2
pha			2																						
bha														1											

Final Position

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	1	3	4	6	7	8	9	10	11	12
ligature p	7	1	2	1	2				1				1												
virāma p	18	4	7	21	8	6	3	9	5	3	12	4	1	3		2			3	6	2				

13.3 Final spellings

In final position two graphs are employed, *pa* and *pa*. Their distribution is strictly dependent on an orthographic phenomenon which I have been calling *virāma specialization*.[63] The principle operates as follows: Whenever a consonant phoneme is to be written in virāma (not to be confused with final position![64]) a new or nonindic radical is to be used if an appropriate one exists. In the present case, *pa* is the virāma specialized sign. The scribes had a clear preference to use virāma (117 ×) rather than ligature (15 ×) to write a final labial plosive, and irregardless of the initial system the graph in virāma is always *pa*. The seven manuscripts which have ligature finals (A, B, C, D, E, I, N) feature always *pa* in this position. In these same documents *pa* is also a regular initial radical; no contrast is being marked. Again unfortunately there are no ligature finals in those manuscripts with predominantly *bha* initials (F?, H, K, 6, 10?, 12?) but there are virāma finals with *pa*. Presumably, if these documents were larger we might find **bh̰* ligature finals, but it is also conceivable that this orthographic school employed just virāma and *pa* for final labial plosives, and never ligature.

13.4 Brāhmī summary

It appears that the predominant spelling system for labial stops used *pa* almost always in nonvirāma, irregardless of linguistic position, and employed *pa* always in virāma. An orthographic rather than a linguistic rule determines the distribution. Other spelling methods are attested only in initial position but these share with the predominant system the use of *pa* in virāma. A larger corpus might have revealed medial and ligature final cases of *ba* and *bha* in those documents having these radicals as regular initials.

The implication of this analysis is that these Turks had one, not two, labial stop phonemes in their speech. And with the consonant shift theory this plosive had to have been the voiced /*b*/. It was written in the majority of the surviving documents with *pa* in virāma and *pa* elsewhere. I hope hereby to have established the voiced value of Brāhmī Turkic *pa*.

63 *Hybrid*, p. 297-298.

64 Not all final consonants occur in virāma, and virāma in the Turkic materials appears not infrequently in medial position, for instance:
oγ̄a `g̰ dı lyā `r, ögdilär, a deverbal noun from *ög-* ‚to praise' (*TTT VIII*, H. 8). I have not systematically studied this latter kind of phenomenon but I suspect that virāma is found only between particular kinds of morphemes and never within a morpheme.

14. Phonemic labial spellings in the Old Turkic scripts

The present theory has the interesting advantage of showing that the majority of Old Turkic scripts represented 100% of the relevant phonological distinctions among the labials. Traditionally these scripts have been regarded as inadequate in this respect. Six writing systems may be identified. With EOT just the Runic was in use but because Runic letters are also found with LOT it is necessary to speak of two separate systems: EOT Runic and LOT Runic. Four additional scripts were employed for LOT: Uygur/Sogdian, Perso-Arabic, Manichaean and Brāhmī.

The EOT dialect had the phonemes /*p*/ and /*b*/ which in Runic writing were clearly and simply rendered with *p* and b^1, b^2 respectively. The LOT dialect had the phonemes /*b*/ and /*v*/. In two writing systems these are represented with a one-to-one, letter-to-phoneme, correspondence. The Uygur/Sogdian script had ‚*p*‘ for /*b*/ and ‚*v*‘ for /*v*/ while the Perso-Arabic had ‚*b*‘ for /*b*/ and ‚*v*‘ for /*v*/. Two other systems spelled the phonemes according to transparent rules. The Manichaean script rendered initial /*b*/ with ‚*b*‘, and noninitial with ‚*p*‘. It may be that this is a partly orthographic rather than strictly linguistic rule; the special initial letter, like the special final letters, helps the reader find word boundaries. /*v*/ is written with ‚*v*‘. The Brāhmī system featured a strictly orthographic distribution of signs with /*b*/ written *pa* in virāma and predominantly *pa* in nonvirāma. /*v*/ is always rendered by *wa*. Only one writing system can possibly be considered as inadequate: the LOT Runic. The orthography was originally designed for EOT and when the phonological change occurred there was a concomitant change in the values of the letters. *p* always stood for /*b*/ as did b^1, b^2 in initial position. A complication possibly existed with noninitial b^1, b^2. While in some dialects this might have always represented /*v*/, it is conceivable that in other dialects in certain clusters[65] a noninitial b^1, b^2 stood for /*b*/. The problems with this script derive from historical circumstance and in no way diminish the high level of linguistic and orthographic competence exhibited by the medieval Inner Asian Turks.

65 See above, 11. Clusters containing labials.