



The Initial Labial Sounds in the Turkish Languages

Gerard Clauson

Bulletin of the School of Oriental and African Studies, University of London, Vol. 24, No. 2.
(1961), pp. 298-306.

Stable URL:

<http://links.jstor.org/sici?sici=0041-977X%281961%2924%3A2%3C298%3ATILSIT%3E2.0.CO%3B2-%23>

Bulletin of the School of Oriental and African Studies, University of London is currently published by School of Oriental and African Studies.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/soas.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

THE INITIAL LABIAL SOUNDS IN THE TURKISH LANGUAGES

By GERARD CLAUSON

I PROPOSE in this paper to discuss only pure Turkish words, that is I specifically exclude all words which were borrowed from some other language at one time or another, and what I shall try to establish is one segment of the phonetic make-up of the oldest form of the Turkish language which we can reconstruct, that is the Turkish spoken appreciably earlier than the eighth century A.D., the date of the earliest substantial specimens of the language which still exist.

One of the arguments on which the Altaic theory is based is the contention that there was in the 'primitive Altaic language' an initial unvoiced labial plosive **p-**, which survives as **p-**, **f-**, or **h-** in some Tungus languages and became **h-** in early (thirteenth-fourteenth century) Mongolian and *zero* in most later stages of that language and in Turkish. This argument is supported by various dubious etymologies, some undoubtedly spurious, which I discuss in greater detail in the Excursus at the end of this paper. The most convenient and recent exposition of this argument is contained in Dr. Pentti Aalto's paper 'On the Altaic initial *p-**' in the *Central Asiatic Journal*, 1, 1, 1955, which begins with the words, 'neither the Turkic nor the Mongolian languages possess an initial **p-** in old original words'.

In a series of articles ending with 'The Turkish elements in 14th century Mongolian', *Central Asiatic Journal*, v, 4, 1960, I have stated my reasons for not believing in the Altaic theory, and I need not enlarge on that subject here; all I want to do is to examine the validity of the statement that the only initial labial plosive which ever existed in Turkish was the voiced plosive **b-**.

Before discussing the plosives, however, I should like first to dispose of the other initial labials. I will begin with the initial labial nasal **m-**.

A good many words with initial **m-** are found in all Turkish languages at all periods except for some early (eighth and immediate following centuries) dialects and some of the so-called Oğuz languages, the most important survivors of which are Azeri, Osmanli (together with its present-day successor Republican Turkish), and Türkmen.

At first sight it might have been supposed that this initial **m-** was part of the original phonetic make-up of the language, but careful examination of the words in which it occurs shows that this is not so. In fact initial **m-** is found only in

- (1) one or two exclamations like **ma**: 'here you are, take this!';
- (2) one or two onomatopoeics like **me**: 'the sound of bleating', and quasi-onomatopoeics like **meme** 'the female breast';
- (3) words which contain, or once contained, a nasal sound, usually as the second consonant, occasionally later in the word.

I shall say nothing more about the exclamations or the onomatopoeics or quasi-onomatopoeics, since philologically speaking they can hardly be reckoned as genuine words. As regards the third class, however, there is overwhelming evidence that in these words the **m-** is an original plosive nasalized by regressive assimilation to the subsequent nasal. This is shown, for example, by the case history of **ben** 'I'. This word appears as **ben** in the oldest Türkü inscription, the memorial to Toñukuk, datable to the first quarter of the eighth century; in the oldest Uyğur inscription, that at Shine Usu, datable to the third quarter of the same century; and in about 80 per cent of the Yenisei inscriptions which are probably a little later, ninth century or in some cases later still (see L. P. Kyzlasov's article (in Russian) 'A new dating for the inscriptions in the Yenisei script', *Sovetskaya Arkheologiya*, 1960, part 3). In the other Türkü inscriptions and documents, in all the other remains of Uyğur and the cognate dialects, and in the literary texts of the eleventh to thirteenth centuries the spelling is consistently **men**, but Kāşğari in the *Dīwānu'l-luġāti'l-Turk*, written in the third quarter of the eleventh century, says¹ that the Oğuz, Kıpçak, and Suwārīn (the last a tribe related to the Volga Bulgars) converted every initial **m-** into **b-** and that thus these tribes said **ben**, while all the other Turks said **men**. Ibn Muhanna, who compiled, probably early in the fourteenth century, an Arabic-Turkish vocabulary, says² in a discussion of the dialectic differences between the Turkish of Turkestan and that of 'our country' (probably Iraq and Azerbaijan), that in the former the word used was **men** and in the latter **ben**. In the thirteenth to fifteenth century Arabic vocabularies of Kıpçak, which include some Türkmen material, the form shown is consistently **men** but Abū Ḥayyān, writing early in the fourteenth century, says³ that the Türkmen form was **ben**. In Anatolia it has been consistently **ben** from the thirteenth century to the present day, but oddly enough Azeri and Türkmen have changed sides sometime during the intervening centuries and both now use **men**, not **ben**.

One or two words are particularly interesting, since at first sight they seem to have an initial **m-** without any nasal letter in the word. The explanation is that the retrospective assimilation was caused not, as is normally the case, by dental **-n-** or guttural **-ŋ-** but by the rare palatal **-ñ-** sound which disappeared almost completely in the ninth or tenth century. One of several words for 'animal dung' is **mayak**. It is found in eighth century Uyğur, chiefly in the medical texts,⁴ in eleventh century Xākānī (Kāşğari),⁵ once in thirteenth

¹ B. Atalay, *Divanü lûgat-it-Türk tercümesi*, Ankara, 1939 (hereafter quoted as 'Atalay'), I, 31 (middle); I, 339.

² *Kitāb ḥilyati'l-insān wa ḥalbatī'l-lisān*, edited by Kilisli Mu'allim Rif'at, İstanbul, 1921, p. 80 (top).

³ *Kitāb al-idrāk li-lisān al-Atrāk*, edited by Ahmet Caferoğlu, İstanbul, 1931, Arabic, p. 36, 2nd column.

⁴ F. W. K. Muller, 'Uigurica', *AKPAW* (Berlin), 1908, p. 29, l. 6. G. R. Rachmati, 'Zur Heilkunde der Uiguren', *SPAW*, 1930 (hereafter quoted as 'Rachmati'), p. 23 (index).

⁵ Atalay, III, 167 (bottom).

century Kıpçak¹ and in one modern language, that of Chinese Turkestan, commonly called Türki in the West and Neo-Uyğur in the Soviet Union and China. Fortunately Kāşğarī, without linking the two, also records² an Oğuz word for 'dung', **baynak**. Both **mayak** and **baynak** are later forms of the original word ***bañak**.

The 'female maral deer' is called **muyğak** in Uyğur,³ in Xākānī (Kāşğarī⁴ and the *Kutadğū bilig*⁵) and in one or two modern languages of the North-Eastern (or South Siberian) group like Teleut and Khakas. Fortunately the Vienna MS of the *Kutadğū bilig* spells the word not **muyğak** but **munğak**. As is well known, this MS is a transcription in the Mongolian Official Alphabet of a text in Arabic script which was very inadequately pointed, and it is full of misspellings. For example **yayig** 'fickle' is more often spelt **tayig**, **yatig**, or even **tatig**. It is quite possible that **munğak** is merely another similar misspelling; but it may represent a genuine phonetic tradition, since the word was no doubt originally ***buñğak**.

The word for 'brain' appears in a most astonishingly wide range of spellings. The earliest occurrences are in an Uyğur medical text⁶ where it is spelt **méyi**. The Xākānī (Kāşğarī⁷ and *Kutadğū bilig*⁸) form is **meçe/méni**. The fifteenth century Çağatay forms are **meye** and **meyin**⁹; the thirteenth to fifteenth century Arabic vocabularies¹⁰ give **meyi**, **meyini**, **beyin**, **beyni**, and **beyini**, the first two obviously Kıpçak, the last three Türkmen. Modern forms range from **me:** and **mi:** in the North-Eastern group to **beyni/beyin** in Osmanli/Republican Turkish. The word must originally have been ***béñi**.

Thus it is clear that in the earliest reconstructable form of Turkish there was no labial nasal initial **m-**; in this respect Turkish was completely different from Mongolian, which contains numerous words with initial **m-** followed by all sorts of consonants in the second place.

There are, in a very limited number of modern Turkish languages, a few words which begin with the voiced labial (bi-labial rather than denti-labial?) fricative **v-**, or even the bi-labial semi-vowel **w-**, instead of the **b-** which occurs in other languages. The initial **w-**, which seems to be peculiar to Kırım (in this and many other respects a South-Western language) is certainly a secondary

¹ M. T. Houtsma (ed.), *Ein türkisch-arabisches Glossar*, Leiden, 1894, Arabic, p. 15, l. 5.

² Atalay, III, 175 (middle).

³ A. von Le Coq, 'Türkische Manichaica aus Chotscho I', *AKPAW*, 1911, p. 35, l. 5. W. Radloff, *Uigurische Sprachdenkmäler*, Leningrad, 1928, 102c.3 and 5.

⁴ Atalay, III, 175 (bottom; misvocalized).

⁵ R. R. Arat, *Kutadğū bilig*, I, *Metin*, İstanbul, 1947, verses 79, 5374.

⁶ Rachmati, lines 63 and 89.

⁷ Atalay, II, p. 299, l. 20.

⁸ Verses 57, 1836.

⁹ *Sanğlax* ('E. J. W. Gibb Memorial Series', New Series, xx), London, 1960, facsimile 319r.28; 321v.10.

¹⁰ Houtsma, Arabic, p. 21, l. 12 (**beyin**). Abū Ḥayyān, Arabic, p. 89, 1st column (**meyi**, in Türkmen **beyni**). *Ettuhfet üz-zekiyye fil-lugat-it-Türkiyye*, edited by B. Atalay, İstanbul, 1945, facsimile 4a.13 (**beyini** in margin); 15a.12 (**meyini**; in margin **beyini**).

form of **v-** and so can be left out of account. So far as I am aware the only languages in which initial **v-** occurs are two modern representatives of the Oğuz group, Azeri and Osmanli/Republican Turkish and the only words in which this sound occurs are **var** 'existing', **var-** 'to go', and **ver-** 'to give', together with a few words derived from these basic words. In these words the initial **v-** seems to have existed continuously from the thirteenth century, the earliest period at which these languages are known, down to the present day. It should be added that there is also in these languages another word, **vur-** 'to strike', with initial **v-** but in this case the **v-** is prosthetic and relatively recent, the original form of the word was **ur-**. We have seen in the case of initial **m-** that the Oğuz group has certain very archaic phonetic characteristics, and it is not impossible that initial **v-** in these words is a survival from a very early period, but the evidence is not conclusive. Kāşğarī, who was familiar with eleventh century Oğuz, does not mention the sound as an initial in that language,¹ and there is no mention of it in any other medieval authority. The verdict therefore can only be 'not proven'.

If there was not an initial **v-** in very early Turkish, the labial series differed in this respect from the dental series where, as I have shown in my paper 'The Turkish *y* and related sounds', *Studia Altaica: Festschrift für Nikolaus Poppe*, Wiesbaden, 1957, there is clear evidence that an initial voiced dental fricative, **ḍ-**, existed in pre-eighth century Turkish.

I now come to the initial labial plosive or plosives and the question whether it is true to say that the Turkish language never possessed an unvoiced labial plosive initial **p-**. The solution of this problem is enormously complicated by the inadequacy of most of the alphabets used to write Turkish and the uncertainty regarding the phonetic values of some of the letters employed. There are, however, some fixed points to start from. The so-called 'Runic' alphabet had different letters for **b** and **p**, and in the relatively limited vocabulary of the texts in this script there is in fact no case of an initial **p-**. The Manichæan Syriac alphabet too distinguished between **b** and **p** and here too in the relatively restricted vocabulary of the texts in this script there seems to be no example of initial **p-**. The position is quite different in the case of the Uyğur MSS in the Brahmi alphabet published by Professor von Gabain in *Türkische Turfan-Texte*, VIII, Berlin, 1954. These texts contain as many initial **p**'s as **b**'s and indeed probably more, but close analysis of the texts shows that of the 15 MSS concerned six have only initial **p-** and **ph-** and three only initial **b-** and **bh-**, while the remaining six use both letters indiscriminately, usually with a strong preference for one over the other. What we seem to be faced with here is a partial breakdown of the elaborate phonetic notation of the Brahmi alphabet. It seems likely that by the time that the alphabet was adopted for writing Turkish the scribes were frankly at a loss to know which letters to use for representing

¹ It should be added, by way of negative evidence, that in Atalay, I, 31, the Oğuz, Kıpçak, Suwārīn equivalent of ('Turkish') **men bardum** is given as **ben bardum**.

some Turkish sounds. No firm conclusions therefore can be drawn from the spellings in these MSS.

There is a gap of some centuries between these MSS and the next document in which an alphabet clearly distinguishing between **b** and **p** is employed. It is the *Codex Cumanicus*, which contains a fairly extensive vocabulary of Koman, a Kıpçak dialect spoken in South Russia early in the fourteenth century, and uses a Latin alphabet to write Turkish. There is no case of a Turkish word with initial **p-** in this vocabulary although Persian loan-words containing this sound are so written.¹ The position is the same in the seventeenth century Koman texts written in the Armenian alphabet and published by J. Deny in *L'Arméno-Coman et les 'Ephémérides' de Kamieniec (1604-1613)*, Wiesbaden, 1957, and in the (rather later ?) Karaim texts written in Hebrew characters, which represent another North-Western language.²

Other medieval Turkish texts are nearly all written in Arabic script. There was no unvoiced **p** in Arabic, and although Kāşğari was aware of the existence of this sound, which he called *al-bā'u'l-şulba*, in Turkish,³ he had no means of writing it. When the Arabic alphabet was adapted for writing Persian, a special letter, triply dotted *bā*, was devised to represent **p**, but this letter was used so sporadically, even to write Persian, that it would be unwise to draw any conclusion from the absence of the two extra dots in any particular case. Indeed it would be more sensible to make no attempt to draw any conclusion regarding the existence or non-existence of initial **p-** from the evidence contained in documents written in Arabic script.

If, therefore, any evidence of the original phonetic structure of the Turkish language is to be found, so far as initial labial plosives are concerned, it can only be in modern languages which have been recorded in scripts which clearly distinguish between **b-** and **p-**. Here the position is one of great variety, indeed even greater than would be inferred from the standard dictionaries. Professor N. A. Baskakov's excellent little book *Tyurkskiye yazyki*, Moscow, 1960, gives a wealth of information about the phonetics of many dialects, including, for example (p. 216), such disconcerting facts as that in the Northern dialects of the Mountain Altai language all initial labial plosives become **m-**. Broadly speaking, however, the position is as follows. A few North-Eastern languages, like Khakas, devoice all initial labial plosives and so have only **p-** except in recent loan-words, mainly Russian. The great majority of modern Turkish languages consistently use initial **b-**, and have initial **p-** only in loan-words, a few onomatopoeics or quasi-onomatopoeics, and a mere sprinkling of other words, but these last may be of great interest. Typical examples of such languages are Kazakh, Kirgiz, and Uzbek. Between these two groups come a few languages in which both initials occur. In no language are initial **p-**'s commoner than **b-**'s, but in all of them there are appreciable numbers of initial **p-**'s. These

¹ See the index in K. Grønbech, *Komanisches Wörterbuch*, Copenhagen, 1942.

² See, for example, T. Kowalski, *Karaimische Texte im Dialekt von Troki*, Cracow, 1929.

³ *Atalay*, I, 8 (bottom).

languages include at any rate Türki/Neo-Uyğur in the South-East, Karakalpak in the North-West, and two members of the South-Western or Oğuz group, Azeri and Osmanli/Republican Turkish. Of all these the last, which is much the best documented, is the language which can be studied to the best advantage. We have already seen that the Oğuz group seems to have preserved in some (though by no means all) respects a more archaic phonetic structure even than some eighth century dialects, and if we can find in it clear evidence of the existence of initial **p-** this phenomenon cannot simply be shrugged off as unimportant.

In fact if we examine the relevant parts of the Osmanli/Republican Turkish vocabulary we cannot fail to be struck by the fact that there is a very clear distinction between initial **b-** and **p-**, and that there is an appreciable number of words with initial **p-** which are, so to speak, consistent in time and space, in the sense that the same words have this initial in nineteenth and twentieth century authorities (the evidence from earlier periods is inconclusive), and that, if a basic word has initial **p-**, words derived from it have the same initial. Some of these are onomatopoeics or quasi-onomatopoeics or words derived from them, like **pat** 'a bursting sound'; **patla-** 'to burst with such a sound', and so can hardly be regarded as significant, but others are perfectly ordinary words which appear in other languages with initial **b-** or occasionally **p-**.

Let me give a few examples of basic words, nearly all of which have a number of derivatives, which are consistently spelt with initial **p-**:

- pek** 'hard, firm; very'
- pert-** 'to bruise (something)'
- parmak** 'finger'
- porsuk** 'badger'
- pus** 'mist'
- pus-** 'to lie in wait'
- piş-** 'to ripen, come to maturity'

All these words have a fairly long history, and three of them, **pek**, **pert-**, and **piş-** occur (with initial **b-**), in the surviving documents written in the 'Runic' or Manichæan Syriac alphabets. There are a good many more, but most of them are either peculiar to Osmanli/Republican Turkish or hardly known elsewhere. For example, the only other authority who mentions **perçem** 'a tuft of hair' is Kāşğari, who says,¹ without specifying the quality of the initial sound, that it is the Oğuz form of **beçkem**, a word otherwise unknown.

Some of these words also appear with an initial **p-** in other languages; for example 'badger' is specifically spelt with initial **p-** in Çagatay² and appears in Karakalpak as **porsuk**, but the most interesting group of words is **piş-** with its derivatives. This verb, which in the early period usually had back vowels but in Türkü³ already had front ones, and its derivatives are still used in almost

¹ Atalay, I, 483 (middle).

² *Sanğlax*, facsimile 132v.22.

³ *Irak bitig*, para. LIII in H. N. Orkuñ, *Eski Türk yazıtları*, İstanbul, 1939, II, 87.

all modern languages, and have an initial **p-** even in languages in which this initial is practically unknown. For the reasons stated above not much importance needs to be attached to the spellings **pişrun-** and **pişrunul-** in a text in Brahmi script, nor to the forms **piş-**/**pis-** in the North-Eastern group, but the words are also spelt with initial **p-** in Türki/Neo-Uyğur, Kazakh (but not Kirgız), Uzbek, and three North-Western languages, Kazan Tatar, Karakalpak, and Nogay and, at any rate sometimes, Türkmen, but oddly enough not Azeri. It is difficult to explain this wide range of occurrences of initial **p-**, even in languages in which this sound is very rare, except as reflecting a very old phonetic phenomenon.

Thus it seems to me that a clear case can be made out for the contention that in the earliest form of Turkish which we can reconstruct from the evidence now available there were two initial labial plosives, voiced **b-** which was very common and unvoiced **p-** which was rather rare.

Considering the rarity of the initial **p-** and the difficulty of proving that it was an element in the original phonetic make-up of Turkish, it would be unreasonable to blame Dr. Aalto for asserting categorically, on the authority of the distinguished scholars whom he was quoting, that the sound did not exist, but the fact remains that this categorical assertion is not justified and cannot be used to bolster up the Altaic theory.

The deeper we dig into the prehistory of Turkish, the greater the phonetic variety which is disclosed. My investigation of the Turkish **y** has shown that there were before the eighth century an initial voiced dental fricative **ḡ-** and an initial palatal nasal **ñ-**, both of which had become **y-** by that date. The present investigation has shown that there was an initial unvoiced labial plosive **p-**, which still exists in some places, and possibly, but not certainly, an initial labial fricative **v-** (but not an initial labial nasal **m-**). In all cases the sounds were rare ones in this position, which explains why their existence has not hitherto been realized. It is perhaps not too much to hope that further diligent search may disclose the existence of other initial sounds the existence of which is not at present suspected.

EXCURSUS

The etymologies quoted to illustrate the evolution of the supposed Altaic ***p-**.¹

In a review of Professor Poppe's *Introduction to Mongolian comparative studies*, Helsinki, 1955, in *JRAS*, 1956, 211, I suggested that while there was good evidence for the phonetic relationship *zero-h-f-p*, it was logically and chronologically more satisfactory to suppose that in the Mongolian and Tungus languages the chain started with *zero* and ended with **p** and not vice versa, in other words that what happened was that in certain circumstances an initial vowel became aspirated and that the aspirate in due course became an **f**,

¹ References like (p. 9) are to the pages of Dr. Aalto's paper quoted above.

and the **f a p**. Evidence for the occurrence of all these sound changes can easily be found in Eastern Asia ; for example in Türki/Neo-Uyğur, a language spoken in an area adjacent to Mongolia, there are a few instances of words with initial vowels becoming aspirated (probably quite recently) and all **f**'s in Arabic loan-words have become **p**'s. While admittedly the earliest dates at which words are known to have existed are not necessarily in the same chronological order as the dates at which those words actually came into existence, so that, for example, a thirteenth century Mongolian word may be 'younger', phonetically speaking, than a twentieth century Tungus one, it would at any rate be more consistent with the phonetic aspect of the latter part of the words concerned to arrange the words for 'hammer' (p. 9) in the following chronological order : Mongolian **aluqa**—Oroc **xaluka**—Manchu **folgo**—Goldi **palū**.

If this is right, then of course the supposed derivation of the word from Akkadian *pilaqu* (p. 12) is obviously false, and this is supported by the fact that the word does not exist in Turkish (surely an essential geographical link between Iraq and Mongolia), except as a recent Mongolian loan-word in one or two North-Eastern languages. The attempt (p. 12) to derive Turkish eighth century **baldu**: 'axe' (**baltu**/**balta** is not attested before the fourteenth century) from a supposed Akkadian *paltu* can only be regarded as extremely venturesome, having regard to the vast intervals in time and space between the two words, and in any case goes dead contrary to the theory which predicates an intermediate 'Altaic' **paltu*, of which the **p**- should become *zero* in Turkish.

'To blow' (p. 11) is, and always has been, **ür-** in Turkish, and never **hür-**. The Türki/Neo-Uyğur words **hür-** 'to bark', for **ür-** (a different word from **ür-** 'to blow'), and **hördek** 'duck', for **ördek**, are typical examples of secondary aspiration in that language. There is not, and never has been, a word **hür-** in either meaning in Osmanli, Çağatay, or any other Turkish language.

Professor Ramstedt's attempt (p. 13) to connect Korean with the 'Altaic group' has aroused such general and well-founded scepticism that it is perhaps unnecessary to examine in detail the arguments put forward in support of it, but the case is certainly not strengthened by attempting to link Korean *pai* 'cup' with Turkish **ayak** through Mongolian *ayaγa*, the form which the word assumed, in accordance with a well-known phonetic law, when taken into that language. Nor does there seem to be any justification for connecting Korean *phul* 'grass' with Turkish **öl** 'moist' (not 'grass' which is **ot**), or Korean *phul-* 'to scratch' with Turkish **üz-** 'to tear, pull to pieces' (not 'to scratch' which is **kaşır-**) or either word with Mongolian *uru-* 'to rub, grate, file, polish'.

It is no doubt true that Turkish **bal** 'honey' (p. 13) is a loan-word from some Indo-European language, perhaps not Latin even though *mel* provides the closest phonetic parallel. Kāşğarī, the earliest authority to quote it says (Atalay, III, 156) that it was used by the Oğuz, Kıpçak, and Suwārīn, which confirms the western connexion. It is in fact a striking exemplification of the intolerance in early Turkish of initial **m-**, which, when encountered in loan-words at that period, was converted into **b-**. The word survives as **bal** (in some North-Eastern

languages **pal**) in all modern language groups, which seems to exclude the possibility of any connexion with Korean *mil*.

There are many other etymologies in Dr. Aalto's paper to which I should feel bound, as a Turcologist, to take exception but I do not think that I have passed over any which would make a better case for the existence of an Altaic ***p-** than those which I have discussed above.