

Orthography and Phonemes in Pashto (Afghan)

Herbert Penzl

Journal of the American Oriental Society, Vol. 74, No. 2. (Apr. - Jun., 1954), pp. 74-81.

Stable URL:
http://links.jstor.org/sici?sici=0003-0279%28195404%2F06%2974%3A2%3C74%3AOAPIP%28%3E2.0.CO%3B2-F

Journal of the American Oriental Society is currently published by American Oriental Society.

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/aos.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.

http://www.jstor.org
Sun Feb 10 08:08:13 2008

All auspicious *li* names which occur more than once and in different localities occur more than once in different places of the Kansu region. In the case of the geographical names, three occur in different parts of the same area (Honan). But one occurs once in the Honan and once in the Hopei area; 3 occur once each in Kansu and once each in Honan; one occurs once in Kansu and once this same name reoccurs in Hopei. It is possible that immigrants coming from the central parts of China transferred their local names to their new settlement in Kansu. Population data seem to substantiate this theory,[12] but the material at hand does not suffice to prove this hypothesis.

In conclusion it might be stated that *li* names in medieval China show certain peculiarities which deserve more extensive study, because they may lead to a better understanding of the history and type of Chinese settlement. They appear, however, to yield fewer names of aboriginal tribes than names of *hsiang* (county) or *hsien* (district) and to be more closely connected with the Chinese settlers as far as the Kansu region is concerned.

[12] The classical case of this type has been studied by H. H. Dubs (in his "A Roman Influence upon Chinese Painting," *Classical Philology*, 38 (1943), 17 foot-note). He proved that Roman soldiers, called as men of Li-chien, taken prisoner in 36 B. C., were resettled in Kansu in a settlement with the same name Li-chien. The documents, unfortunately, seem to mention only one Han time citizen of this place, a certain Erh Ts'ang, 30 years old, 7.2 Chinese feet high, dark complexion, a higher army official (*Chü-yen Han-chien*, vol. 3, p. 6b); it seems very unlikely that this man was one of these prisoners.

ORTHOGRAPHY AND PHONEMES IN PASHTO (AFGHAN)

HERBERT PENZL
UNIVERSITY OF MICHIGAN

1. INTRODUCTION

1. ORTHOGRAPHY AND SPELLINGS, the written reflexes of the phonemic systems, are the primary source material in historical linguistics. There are linguists in this country who consider the study of native writing in its relation to native speech of no importance in descriptive linguistics. It is true that the deliberate separation of orthography and language in the elementary teaching of oriental languages has turned out to be a great pedagogical improvement. It is less certain that the cavalier treatment of native orthographic systems by some linguists can be justified; it mostly provided a convenient rationalization for their own "illiteracy." In communication, orthographies constitute written codes for language messages. The relation between language message and its code is always of relevance to a linguist, particularly if the code is analyzable and describable by linguistic techniques.

Mutual interaction between writing and speech, such as spelling-pronunciations, hyperurbanisms, naive spellings, has been observed in the great literary languages of the West. Similar relations should be investigated in languages elsewhere. It is the purpose of this article to examine briefly the correlation between the native orthography and the phonemic systems of Modern Pashto and its dialects. We shall first take up the prosodic features (**4.**), then the Pashto vowel phonemes (**5.**) and the Pashto consonant phonemes (**6.**), and their representation.

The dialectal variation of Pashto is considerable, but only three standard types have emerged: the type of Peshawar; the Eastern type; the Kandahar type. The differences between these three main dialectal types are not clearly expressed by corresponding regional variations of the orthography.[1] We shall find, however, that there is substantial agreement between the consonantal phonemes of the Kandahar dialect (**6.3** f.) and the orthographic system universally used, while the other two main dialectal types and their subdialects do not agree with the orthography. A history of the rise and development of Pashto orthography cannot be written, since specific historical references are lacking and the available Pashto manuscripts are unfortunately neither old nor numerous enough to provide sufficient source material.

[1] Penzl, *JAOS* 71 (1951), 97.

2. The Pashto Alphabet.

2.1. Pashto is written in the Naskh script of the Arabic alphabet both in Afghanistan and in western Pakistan. The Pashto alphabet contains also the Arabic symbols of the Persian alphabet. In addition to these there are eight special Pashto symbols: for the retroflex consonants *rr* (*rree*), *nn* (*nnun*), *dd* (*ddaal*), *tt* (*ttee*) (**6.2**); for the clusters *ts* (*tsee*), *dz* (*dzee*) (**6.3**); for the spirants *ss* (*ssee*), *zz* (*zzee*) (**6.4**).

The Pashto names of the symbols show the corresponding consonant phonemes in initial position: e. g. /b/ in *bee*, /nn/ in *nnun*, etc. Afghan grammarians [2] usually list 30 alphabetic symbols in the following order: *alif bee pee tee ttee dzhim tshee khee tsee dzee* [3] *daal ddaal ree rree zee zhee zzee sin shin ssin* (*ssee*) *ghayn kaaf gaaf laam mim nun nnun waaw hee yee.* They call these symbols "basic sounds" (*aṣli aawaazuna*, Pers. *aawaazhaa i aṣli*) or "true letters" (Pers. *ḥuruf i ṣaḥiḥ*) (A 1 f., R IV. 4-8, H 116), and usually also list the eight special Pashto symbols separately.

2.2. In Afghan grammars the remaining ten Arabic symbols that only appear in Arabic-Persian loan-words are usually given in a separate list: *fee qaaf 'ayn ḥee ṯwee ṩwaad ṣee ẓwaad ẓaal z̄wee.* They are called "Arabic letters" (*'arabi ḥuruf*) or "borrowed sounds" (Pers. *aawaazhaa i musti'aar*) (A 1, R IV. 9 f., H 116). Only the first four Arabic symbols correspond potentially to phonemes (**6.11**): /f/ /q/ /'/ /ḥ/; the others are just additional symbols for /t/ /s/ /z/.

3. A Latin Alphabet for Pashto

3.1. It is inconvenient in our discussion of Arabic orthography to have to use a Latin transcription exclusively. But such a transcription is indispensable in phonemic discussions. The question of the creation of an auxiliary Latin alphabet for Pashto has received some attention in Afghanistan. No alphabet has been officially adopted, but various government agencies have their own pet schemes. Previous foreign studies of Pashto, e. g. by Raverty, Bellew, Trumpp, Roos-Keppel, Lorimer, Gilbertson [4] have used a variety of symbols.

It is necessary to increase the number of Latin symbols, if each phoneme is to be represented by one symbol. The following methods have been adopted: the borrowing of symbols from other alphabets; the use of different type, e. g. of capitals or italics, for different phonemic values; the creation of entirely new special symbols; the modification of existing Latin symbols, e. g. by dots, points, and lines; the use of a combination of Latin letters for a single value. This last-mentioned "cluster"-writing seems typographically most convenient: e. g. *sh nn*, etc.; it is phonemically unambiguous in Pashto, since the language has no /h/-clusters nor geminate clusters.

3.21. The relation of any proposed Latin alphabet to the Arabic orthography of Pashto has to be carefully considered. A mere transliteration of the Arabic alphabet is no solution, of course. If the shapes of the native orthographic symbols reveal the phonemic pattern, the Latin symbols should, if possible, do the same. Our practice of writing retroflex fortis consonants as geminates, e. g. *rr*

---

[2] We shall quote, as in *JAOS* 71, the following Afghan grammatical studies by the initials of their authors:

M = Ṣaaliḥ Muḥammad Khaan, *Passtoo Zheba*, 1st part: Kabul 1316 (1937); 2d part: 1317 (1938) (in Persian);

A = Muḥammad 'Aẓam Ayaazi, *De Passtoo Qawaa'id*; Kabul, 1324 (1945);

R I = Ṣadiqullah Rishtin, *De Passtoo Keli*, 6th vol.; Kabul, 1326 (1947);

R II = *De Passtoo Keli, Lumrray Lyaarssuwunkey*, Kabul 1326 (1947);

R III = *De Passtoo Ishtiqaaquna aw Terkibuna*, Kabul 1327 (1948);

R IV = *Passtoo Graamar dzhuz i awal*, Engl. title: *Grammar of Pashtu, First Part*; Kabul, Qaws 1327 (December 1948) (in Persian);

H = 'Abdul Ḥay Habibi, *De Passtoo Adabiaatu Taarikh*, vol. I; Engl. title: *A History of Pashtu Literature*, Kabul 1325 (1946);

Rah. = Qazi Rahimullah Khan, *The Modern Pushtu Instructor*, vol. I, Peshawar, 1938 (in English).

[3] The sequence of the five symbols that consist of modifications of Arabic *ḥ* varies sometimes: A 1: *dzh kh tsh ts dz*; R III. 3: *dzh tsh kh dz ts*; R IV. 6: *dzh tsh ts dz kh*. R IV. 8 f. would prefer uniform names for all letters: *dzhee* or *dzha* (*dzhe*) for *dzhim*; *shee* or *sha* (*she*) for *shin*, etc.

[4] Cf. Major H. G. Raverty, *A Grammar of the Puk'hto, Pus'hto, or Language of the Afghāns* (London and Calcutta, 1867[3]); Henry W. Bellew, *A Grammar of the Pukkhto or Pukshto Language* (1867); Dr. E. Trumpp, *A Grammar of the Paštō or Language of the Afghāns* (London and Tübingen, 1873); also *ZDMG* 21. 10-155 (1867), 23. 1-133 (1869); Major G. Roos-Keppel, *A Manual of Pushtu* (London, 1901); Major D. L. R. Lorimer, *Pashtu*; *Part* I: *Syntax of Colloquial Pashtu* (Oxford, 1915); Major George W. Gilbertson, *The Pakkhto Idiom: A Dictionary* (1932).

nn tt dd, follows the Pashto symbols, which also express a pattern relation to the corresponding nonretroflex consonants *r n t d*: a loop is added to the nonretroflex symbols, either at the right side (for *rr dd*) or below (for *nn tt*). The symbols *sh zh* agree with the native Pashto symbols in indicating a relation to the other sibilants *s* and *z*.

It is desirable to use the same graphic devices in the same meaning. *h* as a "diacritic on the line" indicates spirantal articulation in *kh* and *gh* (**6.1**). Our geminate writing of vowels indicates long quantity (**5.1**): *aa ee oo* (*ii*) (*uu*). The short vowels *a e i u* represent, except for *e*, the usual "Latin" values. We only use diacritics as an optional transliteration feature: e. g. in *s̲ z̲ ẓ*, etc.

3.22. Since there is no established Latin orthography for Pashto, the choice of the symbols will sometimes be determined by the purpose of their use. In teaching Pashto to American students at the Institute of Languages and Linguistics of Georgetown University[5] it was found convenient to use these symbols to suggest the proper pronunciation: *aa* but *ē* (not *ee*) *ō* (not *oo*); *L* (not *rr*) *N* (for *nn*) (**6.21**); *hh* (not *kh*) *R* (not *gh*, for the voiced velar spirant with its uvular trill variants).[6]

3.3. The lack of an established Latin orthography results often also in the arbitrary introduction of symbols with the values of the English or French conventional orthographies in Pashto names. Among postal cancellations of Afghan mail we find the spelling *Kaboul* with French *ou*.[7] The name of the language Pashto has often been spelled *Pushto* or *Pushtu* (with the first *u* as in *Punjab*) by writers using English; this in turn was misunderstood by writers using French or German and spelled: *Pouchtou, Poushtoo, Puschtu*, etc.

The Russian scholar E. E. Bertel's who used Latin symbols for Pashto values in a previous publication introduced in Zudin's recent dictionary an alphabet consisting of modified Cyrillic characters.[8]

4. NONSEGMENTAL PHONEMES AND ORTHOGRAPHY.

4.1. A great deal more work needs to be done on stress, pitch, juncture in Pashto. Kandahar Pashto has three stress phonemes: loud (3), medium (2), weak (1). *'pooriwa'hem* (2-1-1-3) '(I) push' contrasts with *'pooriwa'hem* '(that I) push' (3-1-1-2); *(be)'kssee'nem* (2-3) 'I'll be sitting,' contrasts with *(be)'ksseenem* (3-1) 'I'll sit.' Kandahar Pashto has three pitch phonemes: high (3), medium (2), low (1). *nen dzhu'ma de?* (pitches 1-1-3-2) 'Is today Friday?' contrasts with *nen dzhu'ma de.* (pitches 1-1-2-1) 'Today is Friday.'

Stress is never indicated in Pashto writing. Unique and quite isolated is M's practice of indicating final weak and loud stresses by two variants of the *h* sign (26 and passim). Afghan grammarians are, however, aware of the importance of stress distinctions.[9] Intonation and pitch are only indicated by such orthographic signs as question marks, periods, exclamation points, sometimes by commas.

4.2. Word boundaries within utterances are indirectly marked by the difference between initial, medial, and final shapes of most Arabic letters, e. g. of the symbols for *b p t tt s sh ss l kh gh q k m n*. It is not surprising, therefore, that the spaces between words are often very uneven and irregular, particularly in Afghan handwriting. Sometimes we find particles like *ne* 'not,' *pe* 'in, at' written together with following verb or nominal forms: e. g. *pedee shaan* 'in this manner.' This practice seems to be influenced by a parallel procedure in Afghan Persian.

---

[5] This was the first classroom instruction in Pashto in this country (in 1952/1953). Abdul K. Wardakee, a speaker of the Wardak dialect, and Faqir Nabi Alefi, a bilingual speaker from Logar, served as instructors.

[6] In *SIL* 1 (1943), 16.1-4 *R* was used for *gh*, *l* with a crossbar for *rr*; in *Word* 6 (1950) 70-3: *z̲h* for *zz*, *s̲h* for *ss*; in *ZDMG* 102 (1952), 52-61: *r ṇ ḍ ṭ ẓ̌ ṣ̌* for *rr nn dd tt zz ss*; *ġ* for *gh*.

[7] The map in *Report of the Mission to Afghanistan* (1952) 6 shows odd place-name spellings: *Kabul* (city), *Kaboul* (province), *Mushriqi, Mazar i cherif*, etc.

[8] Cf. *Sovetskoe jazykoznanie* 1 (Leningrad, 1935), 173 ff.; P. B. Zudin, *Kratkii Afgansko-russkii Slovar'* (Moscow, 1950).

[9] R I. 2 f., IV. 12 f. and passim, quotes word-pairs where the position of the stress (*de aawaaz fishaar, khatsh, khadzh, aksint*) determines the meaning: *'fat̲ha* 'victory,' *fat'ḫa* 'vowel *a*' (86). R III. 65 noted the difference in stress between *pree'ssood* '(he) was being left' and *'preessood* '(he) was left.' R IV. 13 differentiates between loud stress (Pers. *khadzh i shadid*) and weak stress (*khadzh i khafif*). He is aware of the connection between the position of stress and the variation between *a* and *e*, *ay* and *ey*, *ee* and *i*, *oo* and *u* in the speech of 'some Afghans.' (R IV. 13, fn. 1, 64 f.)

5. VOWELS AND ORTHOGRAPHY.

5.1. In Kandahar Pashto we find the short vowel phonemes *a* (low front) *e* (mid central) *i* (raised mid front) *u* (raised mid back rounded) and the long vowel phonemes *aa* (low front) *ee* (mid front) *oo* (mid back rounded). *ii* (high front) and *uu* (high back rounded) occur only as elegant or formal pronunciations of Persian loanwords: e. g. *ra'iis* (for *ra'is*) 'president,' *tanuur* (for *tanur*) 'oven.' The following diphthongs occur: *ay ey aw* (*ew*) *ei aaw aay*.

In Afghanistan the traditional Arabic diacritics for the short vowels (*ḥarakaat*) are ordinarily neither written nor printed. Pashto grammarians recognize the existence of four short vowels: *a* (*zwar, zabar, fatḥa*), *i* (*zeer, kasr*), *u* (*peess, ẓamma*) and the Pashto *e* (*zwarakay* 'little *zwar*'). A short superior level line, which was apparently derived from the short superior slanted line for *a*, is the diacritic that was created for *e*; like other short vowel diacritics it is ordinarily never used. M uses it in his grammar to demonstrate the contrast between some forms with medial /a/ and /e/.[10]

The consonant symbols *alif h w y* are also used to render short vowels, long vowels, and diphthongs (**5.4** below). Afghan grammarians refer to them, therefore, as "auxiliary letters" (*imdaadi ḥuruf*, Pers. *ḥuruf i imdaadi*) (R I. 3, III. 3, IV. 15 ff., H 112 f.).

5.21. /a/ is initially written by *alif*, e. g. in *anaa* 'grandmother.' /e/ does not seem to occur in this position. Neither /a/ nor /e/ is indicated in medial position, e. g. in *malakh*, pl. *malekh* 'locust.' Finally both are written as *h*: e. g. in *yawa* (fem.), *yawe* (obl. masc.) 'one'; Afghan grammarians often analyze /a/ and /e/ on the basis of the orthography as "two types of *h*."[11]

/e/ used to be occasionally indicated by the *hamza* sign.[12] This practice was rejected by the orthography conference, which met in the summer of 1948 in Kabul. The *hamza* sign is still written with the *y* symbol to indicate final /ey/ as in *kawey* 'you do' (2nd person plural). (A 2, Note; R IV. 16)[13]

5.22. /i/ and /u/ may be indicated by *alif* initially. Medially and finally there is an increasing tendency to write the symbols *y* and *w* for them in most words, even if no elegant /ii/ or /uu/ pronunciation (**5.1**) exists there. The orthography conference of 1948 recommended, e. g., the spelling of *dzini* 'some, several' with medial and final *y*, of *gutta* 'finger' with *w*, of *kutshnay* 'little' with *w*, etc. Afghan grammarians, among them R IV. 12, 15 f., approve of the practice of writing the sign for /y/ or /ii/ (*ma'rufa yee*, Pers. *yaa i ma'ruf* 'known y') instead of the *i* diacritic (*zeer*), and the sign for /w/ or /uu/ (*ma'ruf waaw*, Pers. *waaw i ma'ruf*) instead of the *u* diacritic (*peess*), because it "eliminates doubt." H 112 f.[14] states that the spellings *l r* and *l w r* for *lur* 'daughter' are the same in Pashto.

5.31. The long vowel /aa/ is always clearly expressed by the *alif* symbol, initially by *alif* with the *madd* sign, e. g. *aashnaa* 'friend.' H 113 considers *aa* (*wlaarr alif* 'upright *a*') as consisting of two *a*, similarly R IV. 15.

5.32. /ee/ is perhaps still written by the *y* sign most of the time. Increasingly it is differentiated from /ii/ (/i/) /ay/ and /y/ by two points placed

[10] R IV. 11 f. obviously refers to the *e* diacritic when he defines *e* as 'coming from *a* in a downward direction' (Pers. *az fatḥa ba ṭaraf i paayaan*). W. Lentz, *ZDMG* 91 (1937), 722 expressed doubt whether the sign would ever become generally used.

[11] Penzl, *JAOS* 71 (1951), 98, fn. 3. R IV. 17, Note states that final vowelless *h* (Pers. *haa i saakin*) is not pronounced. He proves IV. 82 f. that substantives with final *h* or *ḥ* or ' behave morphologically in accordance with the fact that the final letters are silent. Rah. 7 speaks of H soft in (*wana*) 'tree' and H hard in (*nikuh*) 'grandfather.'

[12] Rah. still uses the *hamza* for *e*. In the past, Trumpp, *ZDMG* 21 (1867), 33 deplored the infrequent use of *hamza* for /e/ by the Afghans; similarly J. Darmesteter, *Chants Populaires des Afghans* (1888-90) § 48. 2.

[13] R separates this verbal ending from the *ey* occurring in "some conversations" in unstressed final syllables. A says that a *hamza* is written for "masculine *ey*." M 1. 11 writes *dzey* '(you) go' with the level *e* diacritic, but 2. 7 *khwrey* '(you) eat' with *hamza*. Lorimer (1915) calls *ei* (*ey*) "hamza Yé" and transcribes it as (*ai*).

[14] From the point of view of a speaker of the Kandahar dialect, H recognizes three Pashto vowels (*ḥuruf i 'ilat*), each of which has two "states" (*ḥaalaat*) or sounds (*zzaghuna*), a long sound (*uzzd zzagh*) and a short sound (*landd zzagh*): *aa* and *a* (*zwar*); *oo* (*madzhhul waaw*) and *u* (*peess, landd ma'ruf waaw*); *ee* (*madzhhula yaa*) and *i* (*zeer, landda ma'rufa yaa*). Since colloquial Pashto does not have *uu* (*uzzd ma'ruf waaw*) and *ii* (*uzzda ma'rufa yaa*) like Persian, he considers the *uu* and *ii* symbols available for *u* and *i*.

one above the other underneath the *y* symbol[15] instead of two points placed next to each other. In final position, e g. in *kawee* 'you (sing.) do,' a special *ee* symbol, a 'long reversed *y*' (*uzzda tshapa yee*), which is also found in Persian and Urdu orthography, can be used. The orthography conference of 1948 approved of both symbols for final *ee*. The diphthong /ay/, called 'soft *y*' (*mulayina yee*, Pers. *yaa i mulayin*), is written by the *y* symbol, e. g. in *sarray* 'man.'

5.33. /oo/ is written by the *w* symbol and ordinarily cannot be differentiated in writing from /u/ /uu/ /w/ /aw/ /ew/, e. g. in *moor* 'mother.' The diphthong /aw/, as e. g. in *palaw* 'side,' is called 'soft *w*' (*mulayin waaw*, Pers. *waaw i mulayin*) in Pashto grammars. /ew/, as in *kattew* 'cooking-pot,' is very rare.

5.34. The diphthong /ei/ as in *ddooddei* 'food, bread' is called 'strong feminine *y*' or simply 'feminine *y*' (*ṣaqila taaniṣi yee, taaniṣi yee*, Pers. *yaa i ṣaqila i taaniṣi, yaa i taaniṣi*) because of its occurrence as a feminine nominal ending. It is usually written differently from /ey/ and rendered by a special symbol consisting of a modified *y* symbol, i. e. a *y* with a short downwardly slanted line added at its left end.[16]

5.4. We can summarize our account of the writing of vowel-phonemes by the following table, in which *y*' indicates the *y* symbol with *ee* diacritics (**5.32**), *y*'' the modified *y* symbol for /ei/ (**5.34**).

Phonemes	Written in Initial Position	Medial Position	Final Position
/a/	*alif*	—	*h*
/e/	(*alif*)	—	*h*
/i/	*alif*; (*alif* + *y*)	—; *y*	*y*
/u/	*alif*; (*alif* + *w*)	—; *w*	*w*
/ii(i)/	*alif* + *y*	*y*	*y*
/uu(u)/	*alif* + *w*	*w*	*w*
/aa/	*alif* & *madd*	*alif*	*alif*
/ee/	*alif* + *y*' (*y*)	*y*'; *y*	*y*'; *y*; reversed *y*
/oo/	*alif* + *w*	*w*	*w*
/ay/	*alif* + *y*	*y*	*y*
/ey/			*y* & *hamza*
/aw ew/	*alif* + *w*	*w*	*w*
/ei/			*y*''

6. CONSONANTS AND ORTHOGRAPHY.

6.1. The consonant phonemes of Pashto, particularly of the Kandahar dialect, can be listed as follows:

	labial	dental	alveolar	prepalatal	velar	glottal
Semivowels	*w*			*y*		
Liquids			*l r*	*rr*		
Nasals	*m*	*n*		*nn*		
Stops	*p b*	*t d*		*tt dd*	*k g*	*q* '
Spirants	*f*		*s z* *sh zh*	*ss zz*	*kh gh*	*h ḥ*

There are orthographic symbols available for each and every one of the phonemes listed above. In addition to these we find: (1) the symbols *tshee dzhim* for /tsh dzh/, which can be analyzed as clusters of /t/ and /sh/, /d/ and /zh/ respectively, since they do not differ from other existing Pashto clusters. (2) There are Arabic symbols used as multiple representation for /t s z/ which

[15] Cf. M 2. A 2 describes the points as being *sar pe sar* 'head upon head.' The report on the decisions of the orthography conference as printed in Kandahar's daily *Ṭuloo' i Afghaan* of November 21, 1948 stated: "Medial *ee* will be written by two points on top of one another, *i* by two points next to one another." W. Lentz (1937) called the special final *ee* sign, but not the initial and medial two points for /*ee*/ "allgemein anerkannt."

[16] Trumpp, *ZDMG* 21. 36: "Nicht mit ai zu verwechseln ist a-ī.. 'a' auch immer durch ein Hamzah von ī getrennt." W. Lentz (1937) calls the sign for "feminines *ei* (*aī*) allgemein anerkannt."

we transliterate as: *ṯ ṣ ṣ ẓ ẓ z̄* **(2.2; 3.1.)** No special spelling-pronunciations for these symbols can be observed anywhere.

6.11. The following phonemes do not occur in every type of Pashto: the labiodental spirant /f/, the glottalized (pharyngealized) velar stop /q/, the glottal stop /'/, the glottal (pharyngeal) fricative /ḥ/. They constitute an elegant pronunciation type, very often a spelling-pronunciation. In the speech of most Afghans /f/ is often replaced by /p/, /q/ generally by /k/, /'/ by zero or the length of the preceding (or following) vowel, /ḥ/ by /h/ or /kh/: *faayda* or *paayda* 'advantage, gain'; *farq* or *fark* or *park* 'difference'; *Abdul* or (rarely) *'Abdul*; *ba'd* or *baad* 'after'; *fi'l* or *feel* (*peel*) 'verb.' The elegant phonemes /f q ' ḥ/, which show their anomalous status by lacking the voiced counterparts generally found with members of the stop and spirant series, are largely supported by the orthography. The variation between /f/ and /p/, /q/ and /k/ results occasionally in the development of hyperforms: e. g. *urufaa* for *urupaa* 'Europe.'[17]

6.21. Of special interest are those Arabic symbols that are exclusively used in the orthography of Pashto: the symbols for the retroflexed lateral flap /rr/ and its nasalized counterpart /nn/; for the retroflexed prepalatal stops /dd tt/; for the clusters /ts dz/; for /ss zz/.

rr as in *rrund* 'blind,' *sarray* 'man' is written with the *r* symbol modified by the addition of a loop at the right; it is phonetically between *r* and *l*. It is the influence of the orthography that makes Afghan grammarians consistently link /rr/ to /r/, never to /l/.[18]

Two symbols have existed for *nn* as in *manna* 'apple,' *kunn* 'deaf': a simple symbol consisting of the *n* symbol with a loop below, and a compound symbol consisting of the *n* symbol followed by the *rr* symbol.[19] The orthography conference of 1948 rejected the *nrr* symbol in favor of the *nn* symbol. It is due to the shape of the older symbol that the phonetic character of *nn* seems to be clearly recognized by Afghan grammarians.[20] It is more probable that this cluster symbol was originally created as a 'pattern' symbol, i. e. to express the character of /nn/ as a nasal /rr/, than that it is a reflex of the frequent derivation of *nn* from Proto-Iranian *rn*.[21]

6.22. The symbols for the retroflexed prepalatal stops /dd tt/ are the *d* symbol with a loop to the right and the *t* symbol with a loop below respectively. They offer no support for the assumed original introduction of these phonemes through loanwords from Indic dialects, since they express the Pashto pattern only. In older manuscripts, however, *d t* symbols with four superior dots, the Urdu symbols for these sounds, are occasionally found.[22] The phonetic character of the phonemes is suggested to the Afghan grammarians by the current symbols.[23]

---

[17] This writer's name was often pronounced *fintsel* in Afghanistan. R IV. 9 points out that Arabic letters (*ḥuruf*) are replaced by Pashto letters, e. g. *f* by *p*, *q* by *k* or *kh*, ' by *alif*, *ḥ* by *alif* or *kh*, *ṯ* by *t*, *ṣ ṣ* by *s*; *ẓ ẓ z̄* by *z*: thus *mahbub* 'beloved' becomes *maabub* or *makhbub*, *iẓaafat* 'addition' becomes *izaapat*, *qatl* becomes *katl*, etc. He insists, however, that these words be written in their 'original form' (Pers. *ba ṣurat i aṣl*) in the written language (*taḥrir*). He mentions IV. 10. Note also hyperforms, e. g. *f* for *p* in *farheez* 'abstinence' instead of *parheez*, also *q* for *k*.

[18] A 1 f. states that "*rr* sounds like *r* and its sound is a little heavier than *r*" (*daa de* (*r*) *pe shaan awaaz warkawi, leekin awaaz yee ter* (*r*) *lezz tse drund wi*). R IV. 6 calls *rr* "strong *r*" (*raa i ṣaqil*); "in the moment of articulation the tip of the tongue (*sar i zabaan*) touches the middle part of the palate (*ḥisa i wasiṭi i kaam*)." Rah. 1 calls *r* "R soft," *rr* "R hard," but does not even differentiate between the two in his Latin transcription.

[19] Trumpp, *ZDMG* 21 (1867), 28 encountered the compound symbol in many Pashto manuscripts and comments: "was aber als eine inconsequente Schreibweise nicht zu billigen ist, da es ein einfacher, kein zusammengesetzter Mitlaut ist." Darmesteter, *Chants Populaires* (1888), § 45: 'les Afghans eux-mêmes emploient parfois une signe unique *ṇ*.' W. Lentz, *ZDMG* 91 (1937), 722: "An graphischen Abweichungen von dem bisherigen Gebrauch sei angemerkt, dass die ältere Wiedergabe des retroflexen *ṇ* durch die Ligatur *n* plus retroflexes *ṛ* in Afghanistan allgemein verpönt ist." Rah. uses only the compound symbol, which he transliterates by (*nr*).

[20] A 2 mentions both *n* and *rr* as component parts but stresses the nasal character of the sound, which he even calls 'nasal *n*' (*nun i ghuna*). H 116 calls it a sound compounded (*murakab*) of *n* and *rr*. R IV. 8 states that both *n* and *rr* are pronounced simultaneously.

[21] *shkunn* 'porcupine,' Av. *sukurəna*; *kunn* 'deaf,' Av. *karəna*; *paanna* 'leaf,' Av. *parəna*, etc. Cf. W. Geiger, *Etymologie und Lautlehre des Afghānischen* (1893), § 11. 5, G. Morgenstierne, *An Etymological Vocabulary of Pashto* (1927).

[22] B. Dorn, *A Chrestomathy of the Pushtu or Afghan Language* (1847) 387 f. lists several graphic variants of *dd tt rr g*, which he encountered in his manuscript sources.

[23] A 1 f. calls *dd tt* "a little heavier than *d* and *t*."

6.3. The Arabic *ẖ* symbol with three superior dots, which is used for the cluster /ts/, seems to be patterned on the Persian *tshaa* and Arabic *dzhim* symbols for /tsh/ and /dzh/, which consist of Arabic *h* with three dots below and one dot below respectively. The *ts* symbol was formerly used for the cluster /dz/ too; this is still largely true in western Pakistan.[24] In Afghanistan at the present time the *ẖ* symbol with a superior single diacritic, which is often *hamza*, is in general use.[25]

In most Eastern dialects as well as in Peshawar the *ts* and *dz* symbols are pronounced as *s* and *z*; in Laghman as *ts* and *z*; in Wardak and in Kandahar as *ts* and *dz*; in Khost in the Southern Province both as *ts*. The Afghan grammarians reflect in their descriptions the existing dialectal variation.[26] Historically, the clusters must represent a more archaic type than the single consonants. Comparative evidence established a correspondence of Pashto *ts* and *dz* to Avestan *č*: *tsaloor* 'four,' Av. *čaθwārō*; *rwadz* 'day,' Av. *raočah*.

6.4. The symbol for /zz/ in Pashto consists of the Arabic *r* sign with one dot above and one dot below; it patterns with the symbols for /z/ and /zh/, which are written as *r* with one dot above and *r* with three dots above respectively. The Pashto symbol for /ss/ consists of *sin*, the symbol for /s/, with one dot above and one dot below; it thus offers a parallel to the make-up of the *zz* symbol, and patterns with *sin* and with *shin*, the symbol for /sh/, which is written as *sin* with three superior dots. In the dialects of Pashto these *zz* and *ss* symbols correspond to a variety of phonemes: in Peshawar they are simply multiple graphic representations for the /g/ and /kh/ phonemes respectively; in Quetta, Baluchistan for the /zh/ and /sh/ phonemes. In most Eastern dialects the *zz* symbol is also used for /g/, but the *ss* symbol corresponds to a characteristic palatal spirant /kh'/. In Wardak the *zz* sign is mostly written for a palatal spirant /gh'/, the *ss* sign for /kh'/. Only in Kandahar do we find the retroflexed sibilant phonemes /zz/ /ss/, which stand in the relation to nonretroflexed sibilants /z/ /zh/ and /s/ /sh/ that the shapes, names, alphabetic order of the orthographic symbols suggest. Afghan grammarians in their descriptive statements reflect the dialectal variation.[27] Although foreign observers in the past sometimes recognized the retroflex character of *zz* and *ss* in Kandahar, they failed to recognize the presence of four contrasting phonemes /zz zh ss sh/ and their relation to modern orthography.[28] This unique agreement in Kandahar between orthography and phonemic pattern in the two respects where dialects and orthography often clash, namely in the writing of *ts dz* (**6.3**)

R IV. 6 refers to *dd* as 'strong *d*' (Pers. *daal i ġaqil*) and to *tt* as "strong *t*" (*taa i ġaqil*). He adds a correct description of the articulation of *tt*: "In the moment of articulation the lower part of the tongue is brought in contact with the teeth-ridge above and the front part of the palate (*ḫisa i awal i kaam*)." Rah. 1 refers to *d dd t tt* as "D soft, D hard, T soft, T hard."

[24] R III. 3 Note 1 states that the convenient and unambiguous *dz* sign had been recently invented. Roos-Keppel (1901) lists only *ts* called *tsé* or *tzīm* (sic); Lorimer (1915) 10, 13 lists the *ts* symbol with the names *tsé* or *dzé* and the pronunciations [ts s dz z]. Rah. calls the letter *ts* "Sim or Zim."

[25] M used the *ẖ* sign with one point above and one below, but changed in the second volume to the regular symbol. M's original symbol is also used in Zudin's Pashto-Russian dictionary. Trumpp suggested to Afghan copyists in Peshawar that two superior points be used for *dz* and three for *ts*: "Die Afghānen lachten zwar im Anfang darüber, dass ein Franke ihr Alphabet verbessern wolle, aber sie sahen bald die Nützlichkeit der angedeuteten Unterscheidungspuncte ein" (*ZDMG* 21. 23).

[26] R IV. 6 calls *ts* a 'compounded' (*mudzhamu'a*) sound, close to *t* plus *s*. H 118 calls *ts* 'close in sound to a combination of the sounds *t* and *s* (*gaddu aġwaatu te nizdee dey*) or to the German *tsit*.' R IV. 6 describes *dz* as "close to the sound of *z*, somewhat full and heavy." He adds, however, that "some people" pronounce it as a compound sound. H 118 remarks that those who cannot pronounce *dz* substitute *z* for it. As to Latin symbols, cf. Trumpp, *ZDMG* 21. 19, Lentz, *Lateinalphabet* (1937) 9, 13. Morgenstierne, *NTS* 12 (1940), 88-114 used Lentz's *c* (for *ts*), *j* (for *dz*).

[27] A 2 states that *zz ss* sound like *zh sh* respectively, only their sounds are "a little heavier." H 116 is conscious of the dialectal variation; he calls both *zz* and *ss*, therefore, a "compound sound" (*murakab zzagh*), which is pronounced in different ways (*pe ddawl ddawl adaa keezzi*). R IV. 7 observes that "some tribes" pronounce *zz* like *g*, some like "strong *zh*" (*zhaa i ġaqil*); some tribes pronounce *ss* like *kh* and some like "strong *sh*" (*shin i ġaqil*) but "the basic sound (*aawaaz i aṣli*) is between *kh* and *sh*."

[28] Cf. Trumpp, *ZDMG* 21. 25. Most foreign descriptions deal with the Peshawar dialect only. Roos-Keppel (1901), 4 f.: *zz* 'gay' with *g* as in English, *ss* 'khin' as in Scotch *loch*; similarly Lorimer (1915) 8 f.; Rah. (1938): *zz* 'Gē,' *ss* 'khin.' W. Lentz, *Lateinalphabet* 9 suggests a modified Latin *g* for *zz*, a modified *x* for *ss*, although he considered (p. 7) the "Kandeharer Aussprache" to be the standard. Morgenstierne, *NTS* 12. 89, fn. 2 followed Lentz in using these symbols.

and of *zz ss*, seems to justify the conclusion that the old cultural center Kandahar must also have been the cradle of Pashto orthography. This agreement between orthography and phonemes is probably the reason for the traditional great prestige enjoyed by the Kandahar dialectal variety.[29] Comparative evidence appears to support the archaic character of the Kandahar sibilants, since it points to a derivation of Pashto *zz ss* from Proto-Iranian sibilants:[30] e. g. *uss* 'camel,' Avestan *uštra, tsalweesst* 'forty,' Av. *čaθwarəsata, ghwazz* 'ear,' Av. *gaoša, mazz* '(male) sheep,' Av. *maēša.*

7. CONCLUSION

7.1. Our study of the correlation between Pashto orthography and phonemes has revealed some of the typical discrepancies between phonemes and conventional writing-systems. Some phonemes are not adequately differentiated: e. g. *y* is written for /y ii(i) ay/, sometimes for /ey ee/; *w* is written for /w uu(u) oo aw ew/; *alif* is written for all initial short vowels; final *h̲* for /a/ and /e/; medially, short vowels are not written at all. All this results in a interdialectal orthography, where e. g. distributional distinctions involving *a ay* and *e ey*, or *i* and *ee*, *u* and *oo*, or *e* and *i*, *u* need not be indicated.[31] On the other hand, several symbols are also used for one phoneme: *t t̲* for /t/, *z ẓ ẕ z̄* for /z/, *s s̲ ṣ* for /s/.

7.2. Orthographic practice was found to be still fluctuating in some respects. The special *ee* signs (for earlier *y*), *y* and *w* signs for /i/ and /u/ respectively are increasingly used. The special *ei* sign (for earlier *y* with *hamza*), the *dz* symbol (for earlier *ts*), the simple sign for /nn/ seem firmly established now, at least in Afghanistan. There has been obviously a trend in the last decades towards a greater correlation between the orthography and the phonemic system.

7.3. In a country where the majority of the population is still illiterate, knowledge of orthography is a matter of prestige and intellectual pride. The use of the Afghan Persian phonemes /f/ and /q/ by literate speakers is, therefore, often a spelling-pronunciation, sometimes even a hyperurbanism. Occurrences of the phonemes /'/ and /h̲/ are almost always spelling-pronunciations. Elegant /ii/ and /uu/ may also be influenced by the orthography.

7.4. The comparison between the orthography of the consonants and the Kandahar phonemic system shows a remarkable correspondence not only in the number of distinctive units but even in their patterning. Other dialects, however, show mostly multiple graphic representation, e. g. of /s/ by the *ts*, /z/ by the *dz* (or *ts*), of /kh/ by the *ss*, of /g/ by the *zz* symbols. We have concluded from this close correlation between symbols and phonemes that Kandahar was the cradle of Pashto orthography. The character of the modern dialectal variation, the traditional prestige of the Kandahar dialect, the comparative evidence appear to support this important conclusion.

[29] R. Leach, *JASB* 8 (1839), 2: "The Candharee is reckoned the purest dialect." C. E. Biddulph, *Afghan Poetry of the Seventeenth Century* (1890), 1 calls the Northern dialect "rough and harsh," the Southern dialect as spoken in Herat, Kandahar, Quetta "a soft one." G. Morgenstierne, *Report on a Linguistic Mission to Afghanistan* (1924), 10: "Just as in India the Peshawar dialect is predominant, a modified Kandahar dialect is in use among the higher classes in Kabul." W. Lentz, *ZDMG* 91 (1937), 724: "Als eine Art Standardsprache für Afghanistan schienen mir die Gebildeten die Sprache von Kandehar zu betrachten."

[30] J. Darmesteter, *Chants Populaires*, XXXV ff.; W. Geiger, *Etymologie*, § 19.

[31] Penzl, *JAOS* 71 (1951), 97 f.; *ZDMG* 102 (1952), 57 f.

http://www.jstor.org

LINKED CITATIONS

- Page 1 of 1 -

You have printed the following article:

Orthography and Phonemes in Pashto (Afghan)
Herbert Penzl
Journal of the American Oriental Society, Vol. 74, No. 2. (Apr. - Jun., 1954), pp. 74-81.
Stable URL:
http://links.jstor.org/sici?sici=0003-0279%28195404%2F06%2974%3A2%3C74%3AOAPIP%28%3E2.0.CO%3B2-F

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

[Footnotes]

[1] **Afghan Descriptions of the Afghan (Pashto) Verb**
Herbert Penzl
Journal of the American Oriental Society, Vol. 71, No. 2. (Apr. - Jun., 1951), pp. 97-111.
Stable URL:
http://links.jstor.org/sici?sici=0003-0279%28195104%2F06%2971%3A2%3C97%3AADOTA%28%3E2.0.CO%3B2-C

[11] **Afghan Descriptions of the Afghan (Pashto) Verb**
Herbert Penzl
Journal of the American Oriental Society, Vol. 71, No. 2. (Apr. - Jun., 1951), pp. 97-111.
Stable URL:
http://links.jstor.org/sici?sici=0003-0279%28195104%2F06%2971%3A2%3C97%3AADOTA%28%3E2.0.CO%3B2-C

[31] **Afghan Descriptions of the Afghan (Pashto) Verb**
Herbert Penzl
Journal of the American Oriental Society, Vol. 71, No. 2. (Apr. - Jun., 1951), pp. 97-111.
Stable URL:
http://links.jstor.org/sici?sici=0003-0279%28195104%2F06%2971%3A2%3C97%3AADOTA%28%3E2.0.CO%3B2-C

NOTE: *The reference numbering from the original has been maintained in this citation list.*