Some 'Indic' features in Pashto

Michael M. T. Henderson
University of Wisconsin-Madison

1.0   Introduction

Pashto is spoken by over twelve million people, more than half of whom live in Afghanistan. The rest live in Pakistan, with a few outposts in India and Bangla Desh. At times, it appears that there are almost as many varieties of the language as there are families of native speakers, but the main isogloss runs from Peshawar in the northwest to Jaldak, north of Kandahar in the southeast (Grierson 1921: map facing p. 5). North of the isogloss the language is called ['paxto], with a velar fricative or ['paχto] (uvular), and south of it ['pəçto] (palatal fricative) or ['pəṣtu] (retroflex fricative). Since the Kandahar variety, ['pəṣtu], observes more phonological distinctions than the others, it is usually considered the standard by American and Russian linguists, and by its native speakers.[1] The dialectal variations do not affect the arguments in this paper, so I'll use the Kandahar variety, spelling the name *Pashto*.[2] For the sake of brevity, I'll use the term *Hindi* to refer to Hindi-Urdu.

1.1   Earlier investigations

Pashto has several features of interest to typologists, some of which have led to confusion in pioneering descriptions of the language. The features discussed in this paper are: "ergative" verb agreement in perfective aspect transitive verbs; use of distinctive prepositions + postpositions; retention of the category of gender, lost in most varieties of Iranian; unusual initial consonant clusters; and retroflex stops and fricatives contrasting with dentals.[3]

Since any of the above features can be thought, at least at first glance, to be Indo-Aryan, Pashto has been considered considered an Indic, rather than an Iranian, language: "The Pakkhto, like the Hindī, is a dialect of the Sanskrit as regards its grammatical construction, only Persianised in respect to the bulk of the words composing it" (Bellew 1867:ix). Because of the number of Arabic loans, it has also been considered an Iranian-Semitic mixture: "[Pashto is] in all probability derived from the Zend, Pahlavī, and the Hebrew" (Raverty 1864:4).

Darmesteter's monumental work (1890) proved that Pashto is, indeed, an Iranian language, descended from Avestan or another Old Iranian dialect, having separated from Persian before the Pahlavi or Middle Persian period. Grierson cites Darmesteter and states (1921:5) that Pashto is a "Medic," or Non-Persic, or eastern Iranian language. The "Indic-look" features are accounted for by the fact that "it has borrowed largely and freely from North Western India but, in its essence, it is an Erānian tongue" (p. 9; see also Meillet 1922:44-6).

2.0   The features

2.1   Ergativity

Pashto seems to share with Hindi "ergative" constructions, in which the agent in perfective tenses appears in a different case from that of a perfective intransitive or an imperfective (transitive or intransitive) agent:

1.   Pashto   zə xat likəm   3. Pashto   ma xat likha
            'I write a letter'   'I wrote a letter'
2.   Hindi   mɛ̃ tʃɪʈʈi lɪkhta hũ   4. Hindi   mɛ̃ ne tʃɪʈʈi lɪkhi

---

[1] I was given Pashto lessons by Justice Anwar Wardak, who naturally spoke the Wardaki variety, ['pəçtu]. I wanted to learn the higher-prestige Kandahar variety, ['pəṣtu], so our conversations would have seemed strange to another speaker.

[2] I also, with a very few exceptions in short quotes, use IPA symbols to represent the symbols used by the various linguists whose references I cite.

[3] In the non-Kandahar varieties, the retroflex fricatives correspond to palatal fricatives, still contrasting with velar and uvular obstruents.

In (1) and (2), the verb is in imperfective aspect, and agrees with the agent, 'I'. In (3) and (4) it is in perfective aspect and agrees with the patient of the action, 'letter.'

Persian, with which Pashto is most easily compared for typological purposes in Iranian linguistics, no longer shares this feature with Pashto: agents of all verbs in all aspects appear in the nominative or unmarked case, corresponding to Pashto and Hindi direct (vs. oblique) case. Is this evidence that Pashto has borrowed such an important morphosyntactic feature from the Indic languages? Historical investigation shows that this is not the case, and that Persian has *lost* a feature which in Old Indo-Aryan and Old Iranian was a parallel development. Hindi and Pashto have both retained two case-forms for agentive pronouns, substituting for them the dative forms, which have now become "nominative" in function (Sen 1951:119; Regamey 1954:363-366; Geiger 1893:1-5; Matthews 1953:391-408). The western Iranian languages lost the old nominative, and the eastern ones kept it (Emeneau 1965:41-2; Darmesteter 1890:LXXXVIII-XCI.

2.2   CONSONANT CLUSTERS

Pashto has some very unusual consonant cluster in initial position. Persian has none except in the sophisticated speech of people who also know English. Clusters of the type appearing in Hindi appear in Pashto as well. But Pashto has the following non-Hindi (and non-English) initial clusters: bʈ ɣɾ ɣl lɣ ɣʈ ɣw kʈ χp ʂk lɣ lm lw mɾ mz ng (not ŋg) ndʒ nʂ nm nw pʂ pʈ ps pʃ ɾɣ ɾw sχw sʈ ʃn ʃt tl wl zb zd zɣ dzm zg zw ʒw tsk (Lorimer 1915:200; Pattanayak 1966:20). These clusters appear very frequently in the language. Morgenstierne's comment (1940:89) seems appropriate here: "Pashto in its turn...has been subject to a series of violent phonetic changes, affecting as well vowel quality and quantity as simple and compound consonants."

2.3   PREPOSITIONS AND POSTPOSITIONS

Another distinction between Persic and Indic languages is that the former have prepositions while the latter have postpositions. In this respect the Iranian language is the more archaic: the older Indo-Iranian languages had prepositions and case inflection, losing the latter and keeping the former. Indic languages, in contrast, lost the prepositions but kept the case endings, turning them into postpositions (Sen 1951:2). Pashto appears to have retained both:

3. kor   'house'   pə kor kʂe   'in the house'
4. mez   'table'   tər mez lɑndi   'under the table'
5. derʃ kɑl   '30 years'   tər derʃo kɑlo pori   '30 years hence'

It will of course be observed that the last element in each of the above phrases is some sort of substantive; but the head noun of each phrase is inflected, and the postposed elements are semantically different from their occurrences as single units. Note that particles also have this function:

6. dwe badʒe   '2 o'clock'   pər dwo badʒo   'at 2:00'
7. wrustə   'after'   wrustə tər dwo badʒo   'after 2:00'
8. saɾai   'man'   saɾi tə 'to the man'

Penzl (1955:41) calls these particles "a kind of discontinuous preposition." The case inflections associated with the particles are of the same type as those that appear in Hindi, representing direct and oblique cases. Persian, representing the western Iranian languages, does not exhibit direct vs. oblique case marking: only definite objective status is marked, by the particle –rɑ. Thus Pashto seems to have retained both types of nominal function-signals of its precursors.

2.4   GENDER

Pashto has marked gender in nominal and verbal forms, as does Hindi. Persian has no gender marking in nouns, pronouns, or verbs. There is not a great deal to be said about this: Hindi has retained old Indo-Aryan gender distinctions,

and Pashto has retained—while Persian has lost—the Old Iranian ones. The morphemes marking gender in Hindi and Pashto are quite distinct:[4]

9.		Direct			Oblique	
		Sing.	Pl.		Sing.	Pl.
Hindi	m	-a	-o		-e	-õ
	f	-i	-iyā		-i	-iyõ
Pashto	m	∅	-una		-a	-e
	f	-a	-i		-i	-u

This seems to show that the category of gender in Pashto is not, as some have suggested, borrowed from Hindi, but is the retention of a grammatical category lost in Persian.

2.4   RETROFLEX CONSONANTS

The retroflex consonants of Pashto—or, as will be shown below, some of them—present a different problem than do the features discussed so far, all of which (except, possibly, the clusters) can be seen to have perdured in Pashto from its Avestan beginnings, rather than developing later. These consonants are: [ʈ ɖ ɽ ʂ z̤ ɳ]. All have full phonemic status and occur in all positions, except /ɳ/, which occurs initially only in the name of the graph that represents it, /ɳun/. These sounds caused quite a lot of discussion among early describers of Pashto; for example. Trumpp 1873:5-6:

> It must surprize us at the first look, that the Paṣtō alphabet is not possessed of any Aspirates, and in this respect it agrees with the Irānian idioms, but on the other hand it has preserved the full row of the Cerebrals, whereby it closely approaches the Indian Prākrit tongues, yea, it has even preserved a cerebral ṣ (ش), which has long ago disappeared in the Prākrit and the idioms sprang from it.

Trumpp tried valiantly to relate every Pashto retroflex to its Sindhi counterpart, even /ʂ/ and /ɳ/, which do seem phonetically similar to their Sanskrit counterparts. /ɳ/, however, seems to have defeated him (14):

> In the use of this sound the Pashto does not always follow the track of the cognate idioms, but according to its own fancy it has sometimes changed an original cerebral ṇ again to a dental and sometimes an original dental to a cerebral.

Penzl (1955) simply states, "The influence of the Indic languages upon Pashto accounts for the presence of such retroflex phonemes as ɖ, ʈ, perhaps ɽ, ɳ…"

Darmesteter shows that the problem is not insuperable (1890:XIX-XV; and see the Appendix to this paper):[5]

> Every Pashto word which contains a cerebral is borrowed from India … in a certain number of original Pashto words there appears a ɽ ړ , which represents a former cluster rd, rt, in the same way that Persian l represents a former cluster rd … The two sounds—ɽ borrowed from Indic and ɽ derived from rd, rt—do not seem to differ now in pronunciation.

Pattanayak (1966:20) states:

> [Hindi] /ɽ/ is a phoneme carrying an extremely low functional load. Borrowed items creating contrast between [ɖ] and [ɽ] in the intervocalic position are responsible for [its] … phonologisation.

From Darmesteter's list of Pashto reflexes of Avestan consonants (see Appendix) it appears that only /ʈ ɖ/ cannot be traced back to Avestan in some way, and must have been borrowed from Indic. This is not to say that either (1) all instances of these phonemes in Pashto must occur in borrowed Indic *words*, or (2) all examples of /ɽ ɳ/ etc. came from Avestan rather than Indic. Pashto /peɽa/ (a type of Indian sweet) must have been borrowed from Hindi /peɽa/, not

---

[4] This table represents only one class of nouns in each language. For a complete Pashto declension see Penzl (1955:61), and for Hindi, Kachru (1966:41).
[5] I have translated all long quotations from Darmesteter's original French.

descended from something like **paerda*; rather, the capacity of Pashto to borrow Hindi /ɽ/ and /ɖ/ developed from Avestan rd. As a matter of fact, the cases in which Hindi /ɽ/ corresponds to Pashto /ɽ/ or certain to be late borrowings, for the reasons given above in connection with Hindi /ɖ/. It appears that when Darmesteter lays all Pashto retroflexes to Indic loans he is referring only to /ʈ ɖ/ and late /ɽ/.[6]

Morgenstierne has this to say about the subject (1926:12):

Cerebrals are found, not only in Ind. Loanwords, but also in many words of uncertain origin, containing un-Indian sounds like *x* or *z* … Thus Psht., and especially the Peshawar dialect, has been largely Indianized in its phonetic system; but it is worthy of note that it has entirely rejected the [contrastive—MMTH] aspiration of consonants … The Ind. loan-words in Psht. Are generally drawn from modern Hindostani or Lahnda (in contrast with the remarks of Darmesteter … [p. XVI] the latter source is by far the more abundant).

Specimens of Lahnda being unavailable at this time, I have been unable to investigate this; but the remarkably small number of good correspondences between Hindi and Pashto leads me to believe that it is true.

3.0   CONCLUSION

It seems clear that despite its many superficial resemblances to Indic languages, Pashto is an Iranian language (one of the most conservative); and that the only features in it which cannot be explained by direct reference to Avestan are the stops /ʈ/ and /ɖ/. Emenau (1965:30) makes the following observation:

It is impossible … to suggest in any but the most general way how the retroflexes have developed in the Iranian words in which they are found in such languages as Pashto and Yidgha. Bilingualism, involving Indo-Aryan languages, must be the answer, but to historical and social details can be given to elucidate the process.

My own tentative explanation for all the phenomena discussed above—ergativity, use of prepositions with postpositions, gender distinctions, initial consonant clusters, and retroflex consonants—is the following: the Iranian languages had these features, or the seeds of their development, to begin with. In the western or Persic branch, these features, or potential features, died out before the development of modern Farsi/Dari. In the eastern branch, they were retained and augmented by retroflex stops. This retention was no doubt caused by areal factors; that is, contact with the Prakrits from which the modern Indic languages developed (with Emenau, I consider the Dardic languages a branch of Indic). The fact that aspiration was not borrowed is significant, and leads to the following hypothesis concerning the appearance of the retroflex stops: it is probably the case that [ʈ] and [ɖ] were present in Pashto as allophones of /t/ and /d/ in some contexts, and that borrowings from Indic languages caused the phonemicization of these retroflex allophones in the same manner that borrowed English words have caused the phonemicization of Hindi /ɽ/, which is no longer the intervocalic allophone of /ʈ/. Darmesteter (1890:XXI, XXVII) implies that this is also the history of Pashto /g/ and /d/.[7]

---

[6] Darmesteter committed one serious error in recognizing only two varieties of Pashto, and assuming that in the Kandahar variety [ʃ] and [ʂ] were not separate phonemes. He seems to have been led astray by the orthography, in which the grapheme for /ʂ/ ش represents /x/ in non-Kandahar varieties, which has merged with the [x] from other sources. However, on the basis of dialect geography, it would seem that [ʂ] or something like it was present in "Proto-Pashto" and that, in the other varieties it merged with other phonemes; so Darmesteter's list of correspondences still holds.

[7] From the table in the Appendix one sees that Avestan d > Pashto /n l/ and /g/ > /ɣ/. It seems reasonable to suppose that Pashto [d] and [g] were allophones of /t/ and /k/, and later developed full phonemic status through borrowing and/or vowel loss, after which, if my hypothesis is to stand, [ʈ] and [ɖ] were phonemicized.

REFERENCES

Henry Walter Bellew, *A Grammar of the Pukkhto or Pukshto Language* (London, 1867).

James Darmesteter, *Chants populaires des Afghans* (Paris, 1890).

Murray B. Emenau, *India and Historical Grammar* (Annamalainagar: Annamalai University, 1965).

Wm. Geiger, 'Die Passivkonstruction des Präteritums transitiver Verb im Iranischen.' *Festgruß an Rudolf von Roth* (n.p., 1893), 1-5.

George A. Grierson, *Linguistic Survey of India,* X (1921). Reprint (Delhi: Motilal Benarsidass, 1968).

Yamuna Kachru, *An Introduction to Hindi Syntax*. (Urbana: Dept. of Linguistics, University of Illinois, 1968).

D.L.R. Lorimer, *Pashtu, Part I: Syntax of Colloquial Pashtu* (Oxford, 1915).

W. K. Matthews. 'The Ergative Construction in Modern Indo-Aryan,' *Lingua* 3 (1955).

Antoine Meillet, *The Indo-European Dialects* (1922), Tr. Samuel S. Rosenberg (University of Alabama, 1967).

Georg Morgenstierne, *Report on a Linguistic Mission to Afghanistan* (Instituttet for Sammenlignende Kulturforskinning, (1926) Serie CI-2).

_____, 'Archaisms and Innovations in Pashto Morphology,' *Norsk Tidsskrift for Sprogvidenskap* XII (1940), 88-114.

Debiprassana Pattanyak, *A Controlled Historical Reconstruction of Oriya, Assamese, Bengali, and Hindi* (The Hague, 1966).

Herbert Penzl, *A Grammar of Pashto* (Washington: ACLS,1955).

Henry G. Raverty, *A Grammar of the Puḵkhto, Puṣhto, or Language of the Afghans* (London, 1860).

Constantin Regamey, 'A propos de la "construction ergative" en indo-aryan modern,' *Sprachgeschichte und Wortbedeutung, Festschrift Albert Debrunner* (Bern, 1954).

Sukumar Sen, *Comparative Grammar of Middle Indo-Aryan* (Calcutta: Linguistic Society of America, 1951).

Ernst Trumpp, *Grammar of the Paṣtō, or Language of the Afghāns* (London, 1873).