Numerals and Language Contact: Albanian, Slavic, and Romani

Victor A. Friedman
University of Chicago

Key words: borrowing, etymology, Balkan, Bulgarian, Indic, Ukrainian

In this contribution honoring Eric P. Hamp, I shall examine some facts about Albanian, Slavic, and Romani in the histories of their numerals in order to suggest both how they illuminate one another and how they highlight an important variable in the consideration of etymologies, one that Eric himself has always been careful to elucidate, namely that of diffusion versus transmission.

Hamp (1992) gives us the most thorough account of Albanian numerals and their contribution to the reconstruction of Indo-European. He has argued quite rightly that the perfect etymology must be entirely consistent, but at the same time he recognizes the effects of language contact, e.g. he hypothesizes that Mandrica *ǵu* '2' results from the influence of Bulgarian (*dy* > **dju* > **d'u* > *ǵu*) as does the generalization of *tri* for '3' in both genders (also in Ukraine). He likewise notes the opposite effect in Villa Badesa, a Çam-speaking village in Italy, where the generalization of *tri* for both genders apparently serves to distinguish the Arbëresh from the Italian *tre* (cf. Hamp 1992:880), although Italian influence is seen in the vocalism elsewhere in that dialect and also resulted in the replacement of *dh* with *d* in *diét* '10' (Hamp 1992:867-869; cf. also 871-872, where *tri-* for *tre-* in '30' in Sophikò Arvanitika is presumably due to Greek).

As is well known to Albanians living in the Republic of Albania, the Italian numeral *otto* 'eight' is used in place of Albanian *tetë* in contexts when there is noise or some other interference with oral-aural comprehension owing to the similarity of *tetë* with *dhjetë* 'ten' and *pesë* 'five'. This is especially true when dealing with situations such as speaking on the telephone. One might ask why *otto* for 'eight' and not, e.g. *cinque* for 'five' (ten being more marginal, since it already involves second position). A reasonable hypothesis might be that while *cinque* and *pesë* are no less etymologically related than *otto* and *tetë*, nonetheless, from the point of view of the modern languages, *otto* has the advantage of sharing a kind of /t/ with *tetë* whereas the developments of 'five' have left their surface forms less transparently related. This contributes to our understanding

that it is surface forms that are crucial in diffusion, or, as it is put in modern computerese WYSIWYG, pronounced [ˈwɪziwɪg] ('what you see is what you get').

Furthermore, it is worth noting that such usage does not occur in the Albanian of Kosova or Macedonia. This could be explained by the relative absence of Italian influence in these regions as well as its prestige in the Republic of Albania, but Rexhep Ismajli (p.c.) has suggested that the nasality of Geg *pêsë* contributes to a native disambiguation. Be that as it may, the use of Italian *otto* for Albanian *tetë* is a cautionary example of the kinds of problems that can arise in the etymologies of numerals. While the situation is clear to us in the present moment, it is also exemplary of potential problems for etymologists two thousand years hence. At the same time, it is also a contact-influenced feature of Albanian numerals today, and as such, deserves to be taken into account.

Hamp (1975) has shown how Turkic syntax influenced the formation of the numeral '90' in East Slavic. Eric thus combines the rigor of classical Neogrammarian scholarship with sensitivity to history and society. We can note here that the western boundary separating dialects using what Eric has demonstrated is the syntactic Turkism *devjanosto* (or/and its variants) from those with the older Slavic form meaning 'nine tens' begins at the Northwest corner of the current Polish-Ukrainian state border and follows that border exactly until reaching the village of Tel'my, which is dimorphous, i.e. it has both 'nine tens' and *devjanosto*. The one exception is the village of Lubytiv, southeast of Kovel' in Volyn'ska District, which is also dimorphous, perhaps owing to migration. The boundary then follows the administrative division between Volyn'ska and Lviv'ska districts to Ostriv, in the Roven'skyj Region, where, again, both forms occur. The line then runs south along the Lvivs'ka-Ternopil's'ska border to the border between the Zabaz's'kyj and Zoborivs'skyj Regions. Here the border cuts east to Bilozirka (Ternopil's'ka), which is dimorphous. Staryj Oleksinec' and Gindava are to the north (*devjanosto*), Krasnosil'ci and Gaï-za-Rudoju to the south ('nine tens'). At Bilozirka the line moves south and follows the course of the river Zbruč, which also serves as the administrative boundary between the Ternopil's'ka and Xmel'nic'ka Districts to the confluence of the Zbruč and the Dniestr at Ataky, which, being on the right bank of the Dniester is in the Černivec'ka District but is a *devjanosto* village. From here, the border cuts southeast to Biljavynci in Moldova, which is dimorphous, then to Stol'nyčeni, likewise dimorphous, and then south to the River Prut, which becomes the border between the innovation and the archaism. Two villages west of this border, in Černivec'ka—Ringač and Vančynec'—are likewise dimorphous. The Hutsul dialects of Zakarpac'ka and adjacent parts of Hungary and Romania use 'nine tens' in the north and east but are dimporphous in the south and west, with a few purely *devjanosto* villages at the edges of the dimorphous territory (Rieger1996:267, Zakrev's'ka 1988:map 230). It would appear that the reason for the lack of any entry for '90' in Belarusan dialectal atlases is that there is no variation of this sort in Belarusan. Even the Ukrainain dialects in southwesternmost Belarus (Brest district) have the *devjanosto* type. In this respect, '90' is of more dialectological interest than '40' (*sorok*, etc.), for which no lexical variation is recorded for either Ukraine or Belarus. The *devjanosto* versus 'nine tens' isogloss is so sharp in western Ukraine that it suggests an old politi-

cal-cultural divide—Eric has suggested between an economy dependent of fur trading to the east and agriculture to the west—whereas as the dimorphousness that we find precisely in the Hutsul region suggests an accumulation of cultural complexities.

A more recent influence of Turkish on Slavic numerals occurs in the Bulgarian dialects of what is now Greek Thrace and Turkish Asia. As was already recorded by Mladenov and Kodov (1936:103), Bulgarians in some of the villages around Gümülcine (Greek Komotini) and Asia Minor have some or entirely Turkish numerals in their dialects. Later, Kokkas (2004:38, 200-201) gives evidence that the Pomak dialects around Xanthi have Slavic-Turkic doublets for '1' through '4' and then only Turkish for '5' and above. The use of Turkish numerals among Muslims, like the use of Turkish kinship terms in various Balkan languages, is explicable by the high prestige of Turkish in its Ottoman and ex-Ottoman context, but the fact that the Bulgarian-speaking Christian neighbors of Pomaks who then lived in what is now Greek Thrace also had such practices points to a level of multilingual syncretism that challenges the putative homogeneity of nation-states. Of particlar interest in this regard is the fact that fractions are more resistant to substitution (Kokkas 2004:200-201). Matras and Sakel (2007:50-53) posit—on the basis of their broadly cross-linguistic sample—that more complex numerical concepts are more susceptible to contact-induced change; and yet here we have just the opposite. Thus while *ennó~annó, dve, tri, chetri* compete with *bir, iki, üch, dört* as '1, 2, 3, 4', in the expression of fractions we have only the Slavic forms, e.g. 1/3 is only *at trine annóno*, perhaps because of the definiteness. The generalization of feminine/neuter *dve* is also consistent with Turkish influence. As a result of the absence of gender in Turkish, when Turks speak Macedonian or Bulgarian, it is characteristic of their "accent" that they generalize the neuter.

In another of his truly valuable and voluminous contributions to the history of languages, Hamp (1990)—in response to a shockingly ignorant and arrogant article about Romani by an Armenianist in Ohio who shall remain nameless—discussed both the exemplary Neogrammarian regularity of Romani *lautgezetze* and the valuable contributions of Romani to attestations of the history of Indic in the context of its relative conservatism in the preservation of certain ancient Indic distinctions. It is noteworthy that the numerals for 'one' through 'six' as well as 'ten' are remarkably consistent both in the preservation of their Indic origins and in their dialectal distribution. The relative archaism of the /r/ in *trin* '3' has long drawn notice, as has the fact that precisely '7', '8', and '9' were replaced by Greek *efta, ohto, e[j]n(j)a*, presumably at the time of the earliest contact of Romani with Byzantine Greek, given the European dialectal distribution of these numbers in Romani.

The question regarding why precisely these three numerals should have been copied into Romani, presumably in Byzantine times, relates to our observation about Albanian above. The hint comes from Romani '4', for which *catvāri > cattāri > štar*. Thus, in those numerals where short /a/ is followed by long /ā/, we can expect syncope. In the cases of '7' and '8', this would give us *saptā > *spta > *sta*, and *aṣṭā > *šta*, respectively. The surface similarity of *štar, *sta,* and *šta* is such that selecting numerals from a contact language would be motivated by much the same conditions as the use of *otto* for '8' in Albanian. At the same time, it is not unreasonable to suppose that Romani

commercial practices would have encouraged the preservation of the cryptoglossic function of Romani it its broader context, which in turn leads us to wonder whether the replacement of '7' and '8' took place in non-Greek-speaking Byzantium or former Byzantium. (The numeral '9' in both early Romani and Greek would have been so close that simple contact pressure could explain the copying/convegence.) We see hints of differential pressure in Welsh Romani (as well as Anglo-Romani), where the numerals '6' through '9' are *štắr t'ā dūī, trīn t'ā štắr, dūī (~dūvarī) štắr, štắr t'ā panš*, i.e. '4&2, 3&4, 2-(~2x)4, 4&5', respectively (Sampson 1926:351). This unique development for Romani in turn leads us to wonder whether the dialect that became Welsh Romani left the Balkans at a date earlier than that of other Northwest Romani dialects. On the other hand the borrowing of Dolenjski Romani '6'-'9' from Slavic indicates a more ordinary type of contact pressure without cryptoglossic functions, and cryptoglossia itself is more characteristic of the North dialect group.

As can be seen from the foregoing, which builds on insights skillfully elucidated by Eric Hamp, numerals are a locus of both transmitted archaisms and diffused socio-cultural practice. The use of Italian in Albanian helps illuminate the use of Greek in Romani, while Turkic calques and copies illuminate different stages of contact with Slavic. Taken together, they both enrich our understanding of the history of Europe and caution us against mechanistic approaches to language that fail to take social relations into account.

Bibliography

Hamp E. P. (1975). девяносто '90'. In *Russian Linguistics,* 2, 219-22.

Hamp E. P. (1990). The Conservatism and Exemplary Order of Romani Eric P. Hamp University of Chicago In Salo, M. (ed.), *One Hundred Years of Gypsy Studies.* Cheverly, MD: Gypsy Lore Society, 151-154.

Hamp E. P. (1992). Albanian. In Gvozdanović J. (ed.), *The Indo-European Numerals*, Berlin: Mouton de Gruyter, 835-922.

Kokkas N. (2004). *Úchem so pomátsko*. Ksanthi: Politistiko Anaptyksiako Kentro Thrakēs.

Matras Y. and Sakel J. (2007) *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter.

Mladenov S. and Kodov H. 1936. *Bit i ezik na trakijskite i maloazijskite bălgari II, Ezik*. Sofia: Trakijskijat naučen institut.

Reigel, J. (1999). *A Lexical Atlas of the Hutsul Dialects of the Ukrainian Language*. Warswa: Semper.

Sampson J. (1926). *The Dialect of the Gypsies of Wales, Being the Older Form of British Romani Preserved in the Speech of the Clan of Abram Wood*. Oxford: Clarendon Press.

Zakrev's'ka, Z. Ja. (1988). *Atlas Ukraïns'koï Movy*. Kiev: Naukova Dumka.