

THE CLASSIFICATION OF LANGUAGE FAMILIES

Angela Marcantonio

*University of Rome "La Sapienza", Italy.
angmarcan@aol.com*

Abstract: In this paper, we examine the origin of the Uralic languages. We bring together the existing evidence from Lexicon, Morphology, Phonology, human genetics and historical sources. The evidence strongly indicates that these languages do not form a separate family, but instead are "mixed" languages, resulting from relatively recent contact between the Indo-European group - from which a significant part of the lexicon derives - and the languages of Asia and Siberia - with which they share most of the morphology, a significant part of the lexicon, and some key phonological features.

Keywords: Uralic, Finnish, Hungarian, Morphology, Lexicon, Phonology.

1. INTRODUCTION

Most languages of the world have been reliably classified, using the genetic model, as belonging to clear-cut language families. Where language families come into contact, the resulting languages may be classified as 'mixed' and the genetic model does not apply properly.

The languages at the north-eastern boundary of the Indo-European (I-E) area include Finnish, Estonian, Mordvin, Samoyed, and (at the time of its formation) Hungarian. We might expect them to be mixed languages because they are at the boundary of the I-E area. However, they are normally classified in general linguistics (see Comrie 1987) as a well-defined family called 'Finno-Ugric (F-U) / Uralic (U)' (see Hajdú & Domokos, 1978), originated about 8000 / 6000 years ago, as represented in the family-tree diagram at the end of the paper (from Austerlitz, 1987, p. 178). This assumption is widely questioned within the field (see for example Menges, 1968; Sauvageot, 1930; Sinor, 1988.)

In this paper, we bring together the existing evidence from Lexicon, Morphology, Phonology (as well as human genetics and historical sources). The evidence strongly indicates that the U languages do not form a separate family, but instead are “mixed” (see Thomason & Kaufman, 1988) languages, resulting from relatively recent contacts between the I-E group - from which a significant part of the Lexicon derives - and the languages of Asia and Siberia (so-called Altaic (A) and Paleo-Siberian (P-S)) - with which they share most of the Morphology, a significant part of the Lexicon, and some key phonological features.

2. MORPHOLOGY

The morpheme inventory of U is substantially identical with that of A and / or P-S. All the U Local suffixes, and most of the Grammatical ones, are shared with Turkic, Tungusic and /or Mongol (see Aalto, 1969). Furthermore, U shares with A many isomorphic constructions containing etymologically identical morphemes: for ex. a negative verb *e-* (see also Yukagir *ele-* ~ *oi-le*, Finnish imperative / negative *älä* ‘do not’); a past-tense marker *-t*, as in Hungarian and Turkish constructions of the type: *men-t-em-ben* ‘in my going / while I was going’ vs *git-t-im* ‘I went’; an infix (Vowel) *m* (Vowel), to derive non-finite Verbal constructions, in Mongol, Finnish, Turkic (Gerundive), Ostyak (Perfect Participle). The following isomorphic Turkic and Ostyak constructions contain three identical morphemes: the Gerundive / Perfect Participle *-ma-*, the Possessive Suffix and the Locative~Dative vs Lative suffix (respectively) *-a* ~ *-ä*: *tani-ma-ma-m-a* ‘know-not-Ger-my-Loc~Dat. / to my not knowing’ (see Trask, 1996) ; *käs-m-äm-ä* ‘look-Perf.Partic.-my-Lat. / to my looking’ (see Gulya, 1965).

3. LEXICON

The U Lexicon, including the ‘basic lexicon’, is highly mixed - up about to 50 % of the reconstructed U etymologies (see Rédei, 1991) relative to kinship, flora & fauna, body-parts, personal / demonstrative /possessive pronouns etc. can be traced back to I-E, A and / or P-S origin (see Marcantonio, 1996). The Lexicon also contains many cases of ‘multiple /separate borrowings’, where the same word has been borrowed into U separately, from different languages and /or linguistic forms (see Häkkinen, 1990). These phenomena would not be expected in an ordinary contact situation.

There are also multiple indications that the U languages were formed relatively recently (c)

4. PHONOLOGY

Some key phonological U features are shared with A. *Vowel harmony* is considered a basic feature of the U sound-structure, but it is also present in the Turkic languages and in Classical Mongol. *Reduced vowels* are present in Cheremis and the nearby Turkic languages Tatar and Chuvash. Consonant and Vowel Gradation, in one form or the other, are spread all over Eurasia. The reconstruction of the Vocalism of P-U is still quite problematic (see Sammallahti, 1988 and Janhunen 1981), because of several problems, including: a) the difficulty of tracing back to one common source the Vocalism of the major sub-groups: Finno-Permian, Ugric, Samoyed; b) the Hungarian reflexes of the F-U Vocalism are not quite as clear. Furthermore, there are no generally accepted reconstructions of P-Ob-Ugric, P-Ugric or P-F-U sound-structure so far.

5. HISTORICAL AND ARCHAEOLOGICAL EVIDENCE

The historical evidence supports the highly mixed nature of Finnish and Hungarian. The Finns were under the domination of the Goths in the IVth Century (see Kiparsky, 1970). The Hungarians are first mentioned in IXth Century sources, including the Byzantine emperor Kostantinos Porphyrogenetos (*De administrando Imperio*) and other Greek /Arabic sources, where they are consistently described as Turkic tribes. Finally, none of the many attempts to identify a U *Urheimat* based on linguistic palaeontology as well as archaeological findings has proven consistent! Indeed, the few available archaeological findings point to local origins for the various populations (see Branch 1987). The only exception are the Hungarians, who, however, are reported by the above mentioned sources to be living in the area between the North of the Caucasus and the Volga river - no mention or traces of any migration to that area from their supposed home-land around the Ural mountains-.

6. EVIDENCE FROM HUMAN GENETICS

Although genetics does not necessarily correspond to language origin, recent genetic studies (see Sammallahti, 1995) indicate highly mixed populations in the U area, whose genetic components are European and Siberian / Mongoloid (S / M): the U people of Europe, including the Lapps, are mostly European; the U people of Siberia are S / M; the Finns are about 3/4 European and 1/4 S / M; the S / M component in the Hungarians would amount

and from the Mongolians (a ‘barbaric’ people whose language was associated with Finnish under the U-A theory).

8. SOME SUPPORTING DETAILS

8.1 Morphology

Some example of etymologically identical morphemes (in isomorphic constructions)

- negative Verb *e-*: Finn. *älä* ‘do not’, Yukagir *ele-* etc.

- Past Tense / Factitive marker: Hungarian *men-t-em-bem* ‘go-Past-my-locative / in my going / while I was going; Turkish *git-t-im* ‘I went’.

- infinitival morpheme (vowel) *m* (Vowel), used to form subordinate-nominalized sentences: Turkish *tani - ma-ma-m-a* ‘know-not-Gerundive-my-Dative~Locative / to my not knowing’; Ostyak *käs-m-äm-ä* ‘look-Perfect Participle-my-Lative / to my looking’, whereby also the Local endings *-a* and *-ä* respectively are etymologically the same.

8.2 Lexicon

Basic lexicon: some example of U basic lexicon shared with A and /or P-S:

- Finn. *emä* ‘mother’, Hung. *eme*, Yukagir *emei*.
- Finn. *kieli* ‘tongue’, Classical Mongolian *kelen-* ‘to speak’, Tungus *kēlē-* ‘to speak’.
- Finn. *kusi* ‘urine’, Hung. *húgy*, Tungus *ujik* ‘bladder’.
- Vogul *yiv* ‘tree, pine-tree’, Ostyak *yux*, Tungus *jgda* ‘tree’, Turkic (?) *yayač*
- Finn. *hiiri* ‘mouse’, Hung. *egér*, Gold (Tungusic) *siŋgere*
- Finn. *imeä* ‘to suck’, Hung. *emik*, Turkic *em-*, Yukagir *ibi*, Chukchi *emme*

separate / multiple borrowings: Among the several instances of ‘separate / multiple’ borrowings one can quote: Finn. *porsas* ‘pig’ (Syrien *porś*, Votjak *parś*), but Mordvin has an affricate, *purcos*, which is inconsistent and reminds Latin *porcus*. Finn. *mesi* ‘honey’, Hungarian *méz*, (< F-U **mete*), but Cheremis *mü*, Vogul *mag*, Ostyak *mäg*, which instead presuppose a form **meke*. Notice Häkkinen comments (1990, p. 237) to these cases: “Separate borrowing is obviously possible in principle, however it surprises the fact that different, new words meaning ‘honey’ were borrowed separately if it already existed, inherited from Proto-F-U, an I-E word **mete*” ! [the exclamation mark is mine]. Other cases are: **jewä* ‘grain’, Finn. *jyvä* (see below); **mertä* ‘man’, Mordvin *mırde*, Votyak *murt* (see below), but Finnish *marras* (< **martas(e)*), whereby these words might have been borrowed from different forms of the same language.

postpositions: In Ostyak (as well as Samoyed and Cheremis) postpositions are nouns with

attached, like the noun *kim* 'extent': Ostyak *АууА-ем kim-n* 'mouth-my extent-Locative. / up to my mouth'; Hungarian *melle-tt-em* 'chest-Loc.-my / beside me', whereby *mell* meaning 'chest' is used as postposition, *-t-* is an archaic Locative suffix; cmp. however *melle-m-ben* 'chest-my-in / in my chest', whereby the noun is used in its original function.

ethnonyms: See for ex. *udmurt*, the selfdenomination of the Votyaks, whereby the component *ud* is of unknown origin and *murt* (< **mertä*) is of Iranian origin, see Sanskrit *marta-* 'mortal, man'. *Suome* (< *soome*), probably connected to Old-Swedish *some* 'team / group'.

I-E Loan-Words (L-W) : The following considerations apply

- There are only 7 loan-words of I-E origin present in all branches of U (even if such a origin is not properly commented and examined): Finn. *nimi* 'name', *vesi* 'water', *suoni* 'vein', *tuo-* 'bring', *vaski* 'metal, copper', *myy-* 'to sell'; Hung. *mos* 'to wash'; however *nimi*, *vesi*, *tuo-* are also present in Altaic and Yukagir; *vaski* is problematic too, because it is a *Vanderwort* in present all the Euroasiatic area.
- There are some (18 according to Rédei 1986) supposedly Indo-Iranian (I-I) L-Ws which are claimed to have entered at the F-U level; however, given that the specific sound-signs to distinguish between I-E and I-I are scanty, I-E and I-I are in practice used as synonymies. Some examples: Finn. *mesi* 'honey', *mehiläinen* 'bee', *porsas* 'pork', *vasa* 'elk calf', *marras* 'dead', *jyvä* 'grain', *sata* 'hundred'. Some I-I L-Ws are limited to Balto-Finnic (B-F) and Mordvin: Finn. *vasara* 'hammer', *vasa*; and some only to B-F: *taivas* 'sky', *varsa* 'foal / colt'. *Porsas* and *jyvä* are normally classified as F-U (on the arbitrary assumption that, being I-I L-Ws, they must belong to the second oldest layer of the U languages, that is the F-U layer), but they are missing in the Ugric group. Indeed, there is plenty of L-Ws which sound wise are rather archaic, but from the point of view of distribution are quite restricted, like the above mentioned *taivas*, see also B-F *susi* 'wolf' and *tosi* 'truth'.
- L-Ws such as *mesi*, *mehiläinen*, *porsas* and many others are problematic also sound-wise (a part from being cases of 'multiple borrowings'): in *porsas* (< I-E **pork'os*) the I-E vowel **o* of the first syllable and in *mehi-läinen* the I-E **e* of the first syllable have been preserved. therefore from the point of view of vocalism they can be considered archaic.

With regard to the second group, five out of the six listed words, at a closer analysis of the sound shape, turn out to be at least Proto-Iranian (see Korenchy 1972), such as the above mentioned Finn. *sata* (< **śata*) from P-Iranian **śata*; Finn. *sarvi* 'horn' (< **śorwa*), from P-Iranian **śruva*.

- The above mentioned sound changes (merging of the I-E vowel trio and change of I-E palatals into sibilants) are two out of only four phonological criteria available to identify L-Ws of I-E / I-I origin. However (as pointed out by Korenchy 1972, p. 42), "...the sibilant, instead of the I-E palatal, has been preserved for a while in P-Iranian (the change I-E **s* > Iranian *h* being quite late): therefore, if we find a sibilant in F-U and an *s* or *h* (instead of the I-E palatal or **s*) in historical Iranian, we cannot necessarily assume that we are dealing with an I-I L-W, but it could be as well Proto-Iranian"!

REFERENCES

- Aalto, P. (1969). Uralisch und Altaisch. In: *Ural-Altaisch Jahrbücher* (UAJb) 41, pp. 322-334.
- Austerlitz, R. (1987). Uralic languages. In *The major languages of Eastern Europe*, Comrie, B. (Ed.), pp. 177-184.
- Branch, M. (1987). The Finnish language. In *The major languages of Eastern Europe*, Comrie, B. (Ed.), pp. 201-225.
- Comrie, B. (Ed.) (1987) *The major languages of Eastern Europe*. Croom Helm Ltd., London.
- Décsy, G. (1965). *Einführung in die finnisch-ugrische Sprachwissenschaft*, pp 231-241. Wiesbaden.
- Gulya, J. (1965). *Eastern Ostyak Chrestomathy*, p. 124-158. The Hague, Bloomington.
- Hajdú, P. - Domokos, P. (1978). *Uráli nyelvrokonaink*. Tankönyvkiadó, Budapest.
- Häkkinen, K (1990). *Mistä sanat tulevat*, pp. 232-242. Suomalaisen Kirjallisuuden Seura, Helsinki.
- Janhunen, J. (1981). Uralilaisen Kantakielen sanastosta. In *Journal de la société Finno-ougrienne* 77, pp. 219-274. Helsinki.
- Joki, A. (1973). *Uralier und Indogermanen*, p. 362., Mémoires de la Société Finno-Ougrienne 151, Helsinki.
- Kiparsky, V. (1970). Finnougrier und Slaven zu Anfang der historischen Zeit. In *UAJb* 42, pp. 1-8.
- Korenchy, É. (1972). *Iranische Lehnwörter in den obugrischen Sprachen*. Akadémiai Kiadó, Budapest.
- Marcantonio, A. (1996). La "question" Uralica, in *Rivista di Studi Ungheresi* 11, pp. 5-45. Roma 1996.
- Menges, K. (1968). *The turkic languages and people*. Harrassowitz. Wiesbaden.

- Sauvageot, A. (1930). *Recherches sur le vocabulaire des langues ouralo-altaïque*. Société de Linguistique de Paris, Paris.
- Sinor, D. (1988). The problem of the Ural-Altaic relationship, in *The Uralic languages*. Sinor, D. (Ed), pp. 706-742. E. J. Brill, Leiden.
- Thomason, S. - Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press, Berkeley.
- Trask, R. (1996): *Historical Linguistics*, p.126, London
- Veenker, W. (1967): *Die Frage des finnougriſchen Substrats in der Russiſchen Sprache*. Bloomington.

