Esa Itkonen: REMARKS ON THE TYPOLOGY OF THE URALIC LANGUAGES
(Vienna, Sep 26, 2008)

1) The Uralic languages exhibit "considerable typological diversity" (Abondolo 1998:1). So let us decrease the amount of diversity. How? – by raising the level of abstraction. How? – by increasing the time depth, i.e. by going directly to the Uralic protolanguage and then considering a few subsequent developments. In so doing, I will also pay a tribute to my late friend Mikko Korhonen (1936–1991) (cf. Korhonen 1996).

2) The Proto-Uralic (= PU) **verb**: "an inherently stative verb had a present meaning while an inherently punctual verb had a past meaning" (Abondolo 1998: 27). Similarly, e.g. in Hua the least specified verb form expresses "past actions or present states" (Haiman 1980: 136): *ru-e, ri-ne, ri-e* = 'I/you-SG/(s)he took (it)' vs. *bau-e, bai-ne, bai-e* = 'I/you-SG/(s)he stay(s) (here)'. Similarly e.g. Yoruba: *mo/o/ó rà á* = 'I/you-SG/(s)he bought it' vs. *mo/o/ó le* = 'I am / you-SG are / (s)he is hard' (Rowlands 1969: 9, 18).

3) The PU **noun**: Six cases have been assumed traditionally = NOM, ACC, GEN, LAT(ive), LOC, SEP(arative)/ABL. Korhonen (1996: 222-223) reduces this number to five, by assuming the development LAT > DAT > GEN. This is plausible in the light of evidence from e.g. Diyari, (Vulgar) Latin, and colloquial French. The constitution of local case suffixes: LAT = *-x*, LOC = *-x-na*, SEP/ABL = *-x-ta*. It is remarkable that (AL)LAT is the 'unmarked' form, as Korhonen (1996: 149-152) puts it. More frequently, LOC is the unmarked or basic case, which is to be expected on purely logical grounds. For instance, in Hua, Diyari, and Tamil ABL is based on LOC, and in Yagua both (AL)LAT and ABL are based on LOC.

4) How did the PU system of five cases **function**? Typological evidence offers here the possibility of a 'time machine'. The Diyari noun (as opposed to the personal pronoun) has an identical case system, i.e. two grammatical cases and three local cases (cf. Austin 1980, Itkonen 2005a: Ch. I). It possesses a surprising expressive capacity. According to Blake (2001: 156), "in a system of 5 cases, there is a hierarchy such that the last case is an 'elsewhere case', i.e. has no clear-cut function". This may be true of the PU LAT, as shown by its use as a general 'connective' case (cf. Itkonen 1966: 267). On the other hand, none of the Diyari cases qualifies as an 'elsewhere case'.

5) How did the PU case system **develop**? Let us start by considering the **number** of cases. Here the Proto-Indo-European (= PIE) serves as a useful point of comparison. If we disregard VOC(ative), the original 7 cases are preserved in Sanskrit (= NOM, ACC, GEN, DAT, INSTR, LOC, ABL). Otherwise, Lithuanian and Russian = 6 cases (i.e. no ABL), Latin = 5, Cl. Greek and Gothic = 4, French and English = 0. For those familiar only with IE languages, this creates the impression that a full-fledged case system **must decrease** in complexity and be supplanted by periphrastic/adpositional constructions. This view is not contradicted by the emergence of **new** case suffixes or enclitics out of postpositions in Indo-Aryan languages; e.g. Hindi: ERG = *-ne*, DAT = *-ko*, INSTR/ABL = *-se*, perhaps also

LOC = *-mem* (cf. Masica 1991: 230-248). It is also possible to interpret the PREP + ART fusion in the Romance languages as the emergence of a rudimentary sort of **prefixal** case system.

Uralic languages provide here an important corrective, by showing that the history of languages is **less deterministic** than the development of IE case systems may lead one to believe. In the various Uralic languages the number of the cases varies from 3 to 20. In general, there is increase (e.g. Finnish, Estonian, Hungarian), but the number of the cases may also remain the same: "The Samoyed languages have preserved this postulated Proto-Uralic case system quite well" (Korhonen 1996: 199). The same is true e.g. of Mansi. And occasionally the number of the cases may even decrease (as in a Khanty dialect). Uralic (and Dravidian) languages provide counter-evidence to Dixon's (2002) claims that comitative and/or privative/abessive are adnominal cases only (p. 146) and that the lack of double case (or case agreement) is a rarity characteristic of IE languages only (p. 148).

In the present context it is useful to consider the history of Dravidian languages, because it is even **more indeterministic** than the history of Uralic languages. The history of a single language, i.e. Tamil, exhibits extraordinary variation. Proto-Dravidian had at least 4 cases, i.e. NOM, ACC, GEN, DAT (cf. Steever 1998: 20). Ancient Tamil had 7 cases apart from VOC (cf. Lehmann 1994: 38-50): NOM, ACC, GEN, DAT, INSTR, COM, ABL. (INSTR and COM, although expressed by different endings, have also been interpreted as a single case, probably on the model of Sanskrit.) LOC is not yet a case, but a general label for 19 postpositions, each of which also functions as a full lexical noun. Modern Tamil has 8 cases (with a reformed LOC and a totally new ABL). The interesting thing is that in Ancient Tamil the case endings were optional and interchangeable. This state of affairs is asserted by the grammar *Tolkaappiyam*, written some 2000 years ago, and it is confirmed by the extant texts, where both nouns and non-finite verbs appear in their uninflected stem-forms. It is also emphatically asserted by Lehmann (1994: 42, 52-54). Furthermore, I have personally been given confirmations to the same effect by Lehmann and Asko Parpola, two of the leading scholars of Ancient Tamil.This means that, as analyzed in Itkonen (*forthcoming*), a language may simultaneously exhibit two conflicting tendencies (perhaps in two different registers): towards increasingly agglutinatinative structure, on the one hand, and towards isolating structure, on the other. However, the conceptual framework of mainstream typological linguistics is too narrow to accommodate such a phenomenon. Therefore Itkonen (*forthcoming*), written almost 10 years ago, still awaits publication. (But in the meantime, see Itkonen 2000.)

6) Let us next consider the question of the (relative) **self-suffiency** of case systems. It seems that, in order to function properly, every case system needs some additional means of expression. Uralic languages have mainly depended on the use of **postpositions** (cf. Korhonen 1996: 170-174, 208). The same is true e.g. of the Dravidian languages. From the typological point of view, it is interesting to know what **other** options are available.

"Most Australian languages do not have adpositions" (Dixon 2002: 131). It is in agreement with this view that Austin's (1980) grammar of Diyari contains no section devoted to adpositions. This raises two related questions: i) How do languages with no adpositions function? ii) Are there really languages with no adpositions? Let us start by answering these questions for Diyari.

Ad i) Local notions like 'around' or 'into' may be expressed by combining LOC with auxiliary verbs like 'enter' and 'circumvent'. Meanings like 'located in N', '(associated) with N', and 'without N' are expressed by constructions like *N-x* and *N+X*, where *-x* is a derivational morpheme and *X* is an adjective.

Ad ii) In Diyari "uninflected adjectives" (Austin 1980: 106-112) or "modifying nouns" (Dixon 2002: 143) actually seem to function as adverbs that specify the (general) meaning of LOC (e.g. *thati* = 'middle' or *miri* = 'top'). But they can also be seen as incipient forms of adpositions:

(1)	Nhawu	miri	karri-yi	pirta-nhi
	he	above	climb-PRES	tree-LOC

(2) Nhawu karri-yi [pirta miri]-nhi

(3) Nhawu karri-yi pirta miri

In (1), *miri* functions as an adverb. (2) and (3) are synonymous, and in (2) *miri* is an (adjectival) modifier of *pirta*, but in (3), *miri*, which follows the ABS form *pirta-∅,* functions as a prima facie postposition. The same structure occurs e.g. in *Puŋa-∅* ***thati*** *thudu-∅ rdada-yi* = (X) hut-ABS middle fire-ABS make-PRES = '(X) make(s) fire in the middle of the hut'.

Next, let us consider the development that followed PIE: "In the Vedic language ... the particles used as verbal prefixes are also found functioning as [post]positions. But in Classical Sanskrit only two of these **remain** really important, ā ['up to'] and *prati* ['towards'] [, but also *anu* = 'after' and *saha* = 'with' do occur]. ... In addition there is a number of [post]positions of adverbial and nominal origin" (Coulson 1976: 94; emphasis E.I.). Two things should be noticed here. First, there is an **indeterminism** implicit in the fact that 'Vedic > Cl. Sanskrit' entails a decrease in the number of adpositions while 'Cl. Sanskrit > Post-Cl. Sanskrit' entails a corresponding increase. Second, the **adverbial origin** of adpositions (more precisely, prepositions) is still evident in Cl. Greek. The three basic local meanings are expressed as DAT = Location, GEN = Source, ACC = Goal, and if a preposition like *hypo* ('under') or *para* ('near') is added, it agrees with these cases-cum-meanings. Hence, prepositions still do not **govern** their nouns. In Latin the situation has already become more opaque.

If the self-suffiency of a case system is measured by the number of adpositions used, there is a dramatic difference between Sanskrit, on the one hand, and such ancient IE languages as Cl. Greek, Latin, and Gothic. I have checked the number of adpositions on (roughly) the first page of the following works: a) *Bhagavad Gītā, b) Kyrou Anabasis* by

Xenophon, c) *Ab Urbe Condita* by Titus Livius, d) Wulfila's translation of the *Gospel by Mark*: a) 0, b)9 types, 11 tokens, c) 6 types, 11 tokens, d) 9 types, over 20 tokens. Just as importantly, local notions are **never** expressed just by case endings in Cl. Greek, Latin, and Gothic. Thus Blake (2001: 5-6, 32) is wrong to say without qualification that in Latin ACC expresses Goal and ABL expresses Location or Source. Unlike Sanskrit, these other ancient IE languages had **pre**positions.

7) The **origin** and the structure of **adpositions**: Already PU probably had at least such postpositions as **îla-* ('under') and **üli-* ('above'), inflected in the three local cases, "the basic Uralic minimum" (Abondolo 1998: 23). As described by Korhonen (1996: 170), postpositional constructions like *kiven vieressä* ('beside he stone') resulted from the following type of reanalysis: N-GEN **N**-LOC > **N** POSTP (where the boldface represents the main stress) (cf. also Grünthal 2003: Ch. 4). Depending on the language, N-GEN may be replaced by N-∅.

"Most, if not all, African languages ... express prepositional concepts by means of genitive constructions. The result is that the possessed noun phrase assumes the function of an adposition" (Heine & Reh 1984: 101). Swahili: *juu ya mlima* = TOP [OF HILL] > [TOP OF] HILL = PREP N ('on the hill'). Yoruba: *inú ilè* = INSIDE [[∅ = OF] HOUSE] > IN HOUSE = PREP N ('in the house') (Rowlands 1969: 139-141; more precisely: *nínú-u (ilè)* = PREP < *ní inú* = PREP N, where *-u* = GEN of the following N) . The generalization: GEN LOC > N POSTP vs. LOC GEN > PREP N. In Yoruba, as in isolating languages in general, the other strategy is to create adpositions out of serial verbs.

The Tamil ABL *maratt-ile-runtu* combines two distinct strategies: *il* = Ancient Tamil noun ('place', 'house') plus postposition; Modern Tamil LOC = *maratt-ile* = 'in the tree' <'tree-of place'; *iru-ntu* = 'be-past.participle', i.e. 'having been'. When someone falls as one who has been in the tree, the only possible interpretation is that he falls **from** the tree. The other LOC ending *-kitte* (like the identical marker for progressive aspect) goes back to past participle of *kol-* ('to take') (Asher 1984: 111, 163).

(8) The **typological circle** and its implications for Comparative Method (= CR) vs. Internal Reconstruction (= IR): PIE and PU represent different stages of the typological circle: either **agglutination** already combined with **flexion** or incipient agglutination. (Incidentally, the history of Tamil shows that even long periods of agglutination need not be followed by flexion.) Korhonen (1996: 191-194) also sees glimpses of pre-PU **isolating** structure.

According to Korhonen (1983), Saami and Komi have reacted to reductive sound changes in different ways, either creating flexive structure or maintaining agglutinative structure by means of restructuring:

*mene-m *mene-m
*mene-me *mene-j-em

ma**n**am *muna-**m** > mun-**a**
ma**nn**am *mune-**jm** > mun-**i**

Only CM can show that the Saami form *manam* (which participates in a flexive variation) and the (agglutinative) Komi form *muna* descend from the same proto-form *menem*. Here IR would be powerless.

Korhonen (1996: 161) notes that the results of IR are **more abstract** than those of CM. But notice that here 'abstractness', instead of being a virtue, equals 'removed from reality'. Against Fox's (1995: 211) explicit warning, Givón (2000) tries to show that while CM is non-theoretical, IR is theoretical and hence **better** than CM. But he contradicts himself when he speaks (as he must) of the "soft **abductive** underbelly of Comparative Method" (emphasis E.I.). Why? – because 'abductive' equals 'theoretical', and 'soft underbelly' equals here 'falsifiability'. As I have privately pointed out to my friend Tom Givón, if the underbelly of IR is hard, and not soft, then it is not a theory, but a dogma.

9) Finally, let us remind ourselves that the very possibility of typological research is based on interlinguistic **analogy** (cf. Itkonen 2005b: 5-6).

REFERENCES

Abondolo, Daniel. 1998. Introduction. D. Abondolo (ed.): *The Uralic languages.* London: Routledge.

Asher, R.E. 1985. *Tamil*. London: Croom Helm.

Austin, Peter. 1980. *A grammar of Diyari, South Australia*. Cambridge UP.

Blake, Barry. 2001. *Case*, 2nd ed. Cambridge UP.

Coulson, Michael. 1976. *Sanskrit*. Hodder & Stoughton.

Dixon, R.M.W. 2002. *Australian languages*. Cambrifge UP.

Fox, Anthony. 1995. *Linguistic reconstruction*. Oxford UP.

Givón, T. 2000. Internal reconstruction: As theory, as method. S. Gildea (ed.): *Reconstructing grammar.* Amsterdam: Benjamins.

Grünthal, Riho. 2003. *Finnic adpositions and cases in change*. Helsinki: Société Finno-Ougrienne.

Haiman, John. 1980. *Hua: A Papuan language of the Eastern Highlands of New Guinea*. Amsterdam: Benjamins.

Heine, Bernd & Mechtild Reh. 1984. *Grammaticalization and reanalysis in African languages.* Hamburg: Buske.

Itkonen, Erkki. 1966. *Kieli ja sen tutkimus* ['Language and its study', in Finnish]. Helsinki: SKS.

Itkonen, Esa. 2000. Tolkaappiyam: The basic work of ancient Tamil language and culture. *SKY Journal of Linguistics.*

___. 2005a.*Ten non-European languages: An aid to the typologist*. University of Turku: Publications in General Linguistics 9.

___. 2005b. *Analogy as structure and process: Approaches in linguistics, cognitive psychology and philosophy of science*. Amsterdam: Benjamins.

___. *Forthcoming.* A case system with optional and interchangeable endings.

Itkonen, Esa & Anneli Pajunen. 2008. Comments on *World Atlas of Language Structures*. Available on Esa Itkonen's home page.

Korhonen, Mikko. 1983. Reductive phonetic developments as the trigger to typological change: Two examples from the Finno-Ugrian languages. A. Ahlqvist (ed.): *Papers from the 5th international conference on historical linguistics*. Amsterdam: Benjamins. [Reprinted as Korhonen 1996: 207-212]

___. 1996. *Typological and historical studies in language* (edited by Tapani Salminen). Helsinki: Société Finno-Ougrienne.

Lehmann, Thomas. 1994. *Grammatik des Alttamil*. Stuttgart: Steiner.

Masica, Colin. 1991. *The Indo-Aryan languages*. Cambridge UP.

Rowlands, E.C. 1969. *Yoruba.* Hodder & Stoughton.

Steever, Sandford. 1994. Introduction to the Dravidian languages. S. Steever (ed.): *The Dravidian languages.* London: Routledge.