Linguistica Uralica XL 2004 1

REVIEWS * ОБЗОРЫ И РЕЦЕНЗИИ

ANGELA MARCANTONIO (Roma)

WHAT IS THE LINGUISTIC EVIDENCE TO SUPPORT THE URALIC THEORY OR THEORIES?

Abstract. One must always make a clear distinction between things that one has demonstrated with evidence, and things that one has not. Sometimes the latter might be a valuable pointer to further research, however it must be clearly labelled as speculation, because to do otherwise would be misleading to future researchers. If one is to establish a language family within the framework of the conventional Comparative Method, one must begin with a reconstruction of relevant areas of morphology, and the reconstruction of at least the top node, the proto-language. Neither of these elements has to this day been properly implemented in Uralic studies.

Introduction

There are several alternative models to account for the Uralic people and languages. These range from the conventional Uralic theory to the chain model (Pusztay 1997), the bush model (Häkkinen 1983) and the lingua franca model (Künnap 1997; Wiik 2002), as well as others. In order to determine which of these models best describes the true state of affairs, it is clearly necessary to examine the evidence very carefully.

Careful examination of the evidence is the primary objective of my recent book (Marcantonio 2002a). As a result of the detailed review carried out in the book, I showed that the majority of the linguistic evidence on which the conventional paradigm is purportedly based, is simply inconclusive by modern scientific standards. In fact, I argue that the evidence establishes neither the conventional Uralic theory nor the alternative models. The few significant items of evidence that can be identified point to the existence of a series of intersecting isoglosses whose scope is wider than the traditional Uralic area.

This situation could be accounted for by several different models — or it might be the result of a mixture of processes from more than one model, including genetic inheritance, contact, convergence, lingua franca or other processes. The central issue, and the central claim of my book is that, purely on the basis of linguistic evidence and analysis, we cannot tell which of these processes was dominant in creating the situation we observe today. If one is to try to determine which, if any, of these processes was indeed dominant, then evidence from genetics, archaeology and other disciplines outside linguistics must also be taken carefully into consideration. A similar conclusion has been reached, through a different line of reasoning, by scholars such as K. Julku (Itämerensuomi — eurooppalainen maa 1997), A. Künnap (1998) and K. Wiik (2002). Alternatively, perhaps, someone might come to the rescue by proposing an entirely new linguistic approach that has not yet been considered.

These conclusions fly in the face of literally hundreds of years of scientific research in the field, and it is absolutely correct to examine them very critically. Reactions to the book amongst scientists have varied extremely widely. ”Revolutionaries”, such as A. Künnap (2002), K. Julku (2002) and K. Wiik have long supported similar concepts. ”Counter-revolutionaries” tell me that they personally disagree with the conclusions, but they strongly support the opening of a valid scientific debate — I am particularly indebted to J. Janhunen who pro-actively recommended my book for publication on these grounds. P. Kallio and M. de Smit (2002) regard the book as a fundamentally mistaken example of ”voodoo science”. And, finally, there are what I may label ”disinformation theorists”, who tell me they have taken action to prevent students and others in their university from being exposed to my ideas, on the grounds that they represent disinformation.

With this background, it has been a privilege for me to correspond over recent months, actively and on a professional level, with proponents of all these approaches. I believe this has been illuminating for all concerned, and here I seek to share some of the best of the scientific debate and clarification that has taken place on this subject.

J. Janhunen (2001) sets the scene for the debate on the validity of the new approaches. He says it is valid to consider the approaches of the so-called ”revisionists” and ”revolutionaries”, although in his personal view ”there is nothing basically wrong with the conventional paradigm of Uralic comparative studies.”

In addition to recommending how the debate should be conducted, the paper defends the conventional paradigm by focusing on the early works that established the Uralic language family, including H. Paasonen (1912/1913—1916/1917) and M. A. Castrén (1858). Following private discussion relating to these historical papers, a comment of mine on J. Janhunen’s (2001) paper has recently appeared in the ”Diskussion” section of FUF (Marcantonio 2002b), and may be of interest to the reader who wishes to know more of the historical foundation of the conventional paradigm.

Here I seek to set out, as best I am able, the main elements of the approach of those scholars who regard my work as ”voodoo science”. Naturally, it is only possible to focus on what I perceive to be the main issues here. Although I believe that there has been effective private communication on these matters, I must, of course, take full responsibility if I unintentionally misrepresent the views of these scholars.

The Uralic proto-language was first systematically reconstructed by J. Janhunen (1981). Intuitively, this reconstruction seems to constitute compelling evidence in favour of the language family. Given that one is able to go so far as to reconstruct an entire proto-language, surely this must have an important meaning. The paper sets out some 30 diachronic sound-changes, and these are supported by some 90 impeccable etymologies. Clearly, there are more etymologies than rules; so it seems to follow self-evidently that this represents a good scientific model — one that explains many observations with only a small number of adjustable parameters. My book apparently fails to acknowledge this simple truth.

Instead of focusing on the (diachronic) sound-changes that are set out in the Uralic reconstruction, my book simply appears to ignore them. Instead, the book focuses merely on the synchronic correspondences between the languages compared (Finno-Permian and Samoyed). It finds that there are more (synchronic) sound-rules than actual correspondences, and, therefore, it comes to the conclusion that this does not represent a good scientific model because there are more postulates than items of evidence. One might argue that this is sophistry of the highest degree — to ignore the diachronic model that gives the reconstruction its validity, and then to claim not to have found any evidence to support the conventional model!

Worse, the book leaves the job unfinished. For example, after comparing Hungarian with other Uralic and Asiatic languages, it discusses the apparent correlations between Hungarian and Turkic, or between Hungarian and Mongolian, which appear to contradict the conventional approach. It speculates that, if one were forced to choose a simple classification, Hungarian should best be regarded as an Inner Asian language. In a highly unsatisfactorily manner, the book does not propose a specific model to support this statement, thus avoiding the necessity to compare the alternative model with the actual evidence.

It is natural to think that, surely, this author cannot be serious. She fails to recognise self-evident truths about J. Janhunen's reconstruction of the Uralic protolanguage. She does not appear to understand the distinction between diachronic sound-changes and synchronic sound-correspondences, and, apparently, as a result she completely ignores the very diachronic basis of the Comparative Method. And she hides behind speculation rather than proposing proper testable models. It is easy to imagine that such an author must be either malevolent or incompetent, and that the Philological Society and its reviewers must have been subject to some sort of collective dementia in deciding to publish it.

These charges are serious and professionally motivated. Are they valid?

If you live in a University town, you are likely to meet biologists in a social setting. Perhaps their most common dinner-party complaint the world over, is that they cannot publish anything without the statisticians reviewing their work. Many biologists are naturally predisposed to a more intuitive approach. Statistical studies can often produce counter-intuitive results — especially if they purport to show that the data you have been measuring over a long period does not have statistical validity and, therefore, cannot be published in its current form. Nevertheless, biologists have now universally come to accept the ground-rules required by the statisticians.

Perhaps the core of the message in my book is that linguists, too, must adopt the same ground-rules if they are to convince the scientific community of the validity of their results. It is a hard discipline. If you cannot pro-actively demonstrate your results are statistically meaningful, they must either be clearly labelled as mere speculation, or they should not be published at all. This hard discipline must, unfortunately, be adopted generally in the linguistic community.

Synchrony or diachrony?

Statisticians will almost always require that only the measured and attested data can be counted as evidence. Given that there are no ancient written records in the Uralic area, the reconstructed proto-forms cannot be verified directly against the primary sources, and, therefore, they cannot be used directly as evidence.

It is easy to see why a statistician will reject the use of the intermediate outputs of a model, such as proto-forms, as purported evidence in support of the model itself. The circularity of this type of argument is generally recognised in textbooks on linguistic reconstruction. For example, A. Fox (1995 : 63—64) says:

"An inevitable problem here is that of circularity: it is difficult to avoid some version of the vicious circle that results from assuming that forms are cognate because they can be reconstructed with the same proto-phoneme, where the protophoneme is itself the result of assuming that they a r e cognate. For example, on the basis of the pair Latin [*pater*] ~ Gothic [*fadar*] we assume that Latin [*p*] is equivalent to Gothic [*f*] and establish a proto-phoneme **p*. We then conclude that [*pater*] and [*fadar*] are cognate b e c a u s e [*p*] and [*f*] are both derived from the same proto-phoneme [–––] Such circular reasoning is to some extent inevitable with a method that is self-contained [–––] it can lead to serious distortions [–––] "

Any scientific model must be able to make testable predictions, that is, predictions against what we actually observe in the attested languages. This is essentially a synchronic process because of the synchronic nature of the data available in this linguistic area. After the models have been established in this way, it may (possibly) be valid to infer the diachronic processes and/or the historical reconstructions. However, the inference cannot be made the other way round, because the ancient forms are not directly attested.

The founders of the Uralic paradigm, including J. Budenz, appeared to accept these ground-rules. In more recent times D. Abondolo (1998 : 8) puts it thus:

"The historical development of all these families [including Uralic] has been established, by means of the Comparative Method, to a degree of precision that is both predictive and productive. "Predictive" means that given a form X in language Y we can predict, on the basis of regular correspondences and credible courses of development, what the form of its cognate, form Z in related language W, will be."

Whilst I would argue that this programme has not, in fact, been successfully implemented within Uralic, because of the problematic and scanty nature of the evidence, D. Abondolo and I are in complete agreement on the ground-rules that must be adopted and the criteria of proof.

In summary, my book concentrates on the synchronic predictions of the model (i.e. the sound-correspondences) because these can be tested against the actual evidence. The diachronic predictions of the model (i.e. proto-forms and sound-changes) cannot be tested directly against the evidence and so they are of limited use in the context under discussion.

Is the Uralic reconstruction statistically significant?

The first and main part of the paper on the Uralic reconstruction by J. Janhunen (1981) concerns the establishment of the (synchronic) sound-correspondences. The great strength of the paper is that the correspondences and their supporting etymologies are catalogued and numbered in a way that is amenable to quantitative scrutiny. Whilst there is scope for discussion on some details, I understand, from private correspondence, that it is acknowledged the statistical significance of these synchronic rules is at least subject to question, because the number of sound-correspondences is of the same order as the number of regular etymologies.

The main area of useful clarification concerns the (diachronic) sound-changes. My book does not compare these against the evidence for the reasons outlined above. Nevertheless, the diachronic sound-changes might contain fewer adjustable parameters than the synchronic sound-correspondences. If these sound-changes can be combined together into a predictive model that can be compared against the (synchronic) evidence, then, indeed, this would confer more statistical significance to the model. However (as we shall outline below), when we come to examine this quantitatively, we find that the diachronic rules contain more adjustable parameters than synchronic rules, and, therefore, it is not possible to build a model with greater statistical significance by using them.

In the original paper (p. 247—250), there are 11 sound-changes that apply irrespective of context, and 20 sound-changes whose application depends on the context. For example, taking the last rule on p. 247 of the original paper, the sound change /**i̮*/ or /**i*/ > /**∂*/ (in second syllable) occurs in the following contexts: 1) after an original, closed first syllable 2) after a palatalised /*d′*/ 3) after a palatalised /*ń*/ 4) before a syllable-final semi-vowel which, although present in the original, has been dropped over time

Therefore, this rule comprises four sub-rules, so contains, in itself, several adjustable parameters. Further, if one is to construct a predictive model, one would have to add more adjustable parameters in order to resolve this ambiguity of /**i̮*/ or /**i*/. In fact, if one counts the total number of adjustable parameters in all the diachronic sound-rules, this exceeds the number of the synchronic rules.

Even if one were to put aside the reconstruction's lack of statistical significance, there is still a problem. Despite its title as the Uralic reconstruction, this is not really a reconstruction of the Uralic proto-language at all. The key Ugric node has never been reconstructed, and it is simply omitted from the systematic comparison. This should, therefore, be properly labelled the Finno-Permian-Samoyed proto-language, after the two language groups that are systematically compared. There is, in fact, no evidence that this Finno-Permian-Samoyed proto-language can be equated with the Uralic proto-language, as suggested in the original paper: this is a starting-point a s s u m p t i o n and it is based in turn on a chain of several other assumptions (see Marcantonio 2002b : 469). This means that further evidence would be required before the reconstruction of proto-Uralic could be regarded as implemented within the framework of the traditional paradigm, even if it were demonstrated to be statistically significant.

In summary, to this day there remains no systematic reconstruction of either of the key Finno-Ugric or Ugric nodes, and, consequently, of the top Uralic node. This state of affairs is not contentious — it is widely acknowledged, also by the proponents of the conventional paradigm, including P. Sammallahti (1988 : 484), P. Hajdú (1987 : 306—10), J. Janhunen (1998 : 461) and E. Helimski (1984 : 253). Furthermore, it is difficult, if not impossible, to trace back and reconstruct for proto-Uralic major, relevant areas of morphology (such as verbal, grammatical, functional paradigms or sub-sets of paradigms), for the simple reason that most of the (complex) endings of the modern Uralic languages are innovation formed independently in the various languages. This too is a well-known, not contentious fact within the field (compare Korhonen 1996), as it is not contentious that the reconstruction of morphology represents an essential, if not the primary, element in establishing relatedness (see Meillet 1967 : 41; Nichols 1992 : 13).

It is one of the central themes of my book that one must always make a clear distinction between things that one has demonstrated with evidence, and things that one has not. Sometimes the latter might be a valuable pointer to further research, however, it must be clearly labelled as speculation or some equivalent phrase, because to do otherwise would be misleading to future researchers. I believe it is the failure to do so that lies at the heart of the problems with Uralic research today.

Some may regard it as unfortunate that, in setting out my personal views regarding the classification of Hungarian, I am clear that I have not reached adequate standards of proof and, therefore, these views must be described as speculation. It seems to me, personally, that the evidence as examined in my book, as well as in many major works on Hungarian/Altaic correlation (such as Ligeti 1986), strongly point to Hungarian being some type of Inner Asian language. However, I must be clear that I have not produced a full model and tested it against the evidence. This could be a fruitful area for future research — but that would be another book entirely.

Conclusion

If one is to establish a language family within the framework of the conventional Comparative Method, one must begin with a reconstruction of relevant areas of morphology, and the reconstruction of at least the top node, the proto-language. Neither of these elements has to this day been properly implemented in Uralic studies. The scanty reconstructions that have been implemented fail according to the conventional criteria of the Comparative Method: in morphology because it has been impossible to reconstruct sets (or even sub-sets) of grammatical paradigms, in phonology because it has been impossible to reconstruct the key Ugric node, and, therefore, the so-called top Uralic node had to be based only on Finno-Permian and Samoyed. This latter reconstruction also fails the significance criteria that have been accepted in other branches of science, as discussed above. Until these issues have been satisfactorily dealt with, I commend to the reader that it is neither "disinformation" nor "voodoo science" for one to have doubts about the validity of the Uralic language family, at least in the terms it has conventionally been conceived.

REFERENCES

A b o n d o l o, D. 1998, Introduction. — The Uralic Languages, London (Routledge Language Family Descriptions), 1—42.

C a s t r é n, M. A. 1858, Hvar låg det finska folkets vagga? — M. A. Castréns smärre afhandlingar och akademiska dissertationer. Nordiska resor och forskningar 5, Helsingfors, 126—142.

F o x, A. 1995, Linguistic Reconstruction. An Introduction to the Theory and Method, Oxford.

H a j d ú, P. 1987, Die uralischen Sprachen. — P. H a j d ú, P. D o m o k o s, Die Uralischen Sprachen und Literaturen, Hamburg, 21—450.

H ä k k i n e n, K. 1983, Suomen kielen vanhimmasta sanastosta ja sen tutkimisesta. Suomalais-ugrilaisten kielten etymologisen tutkimuksen perusteita ja metodiikkaa, Turku (Publications of the Department of Finnish and General Linguistics 17).

H e l i m s k i, E. 1984, Problems of Phonological Reconstruction in Modern Uralic linguistics. — LU XX, 241—257.

J a n h u n e n, J. 1981, Uralilaisen kantakielen sanastosta. — JSFOu 77, 219—274.

—— 1998, Samoyedic. — The Uralic Languages, London (Routledge Language Family Descriptions), 457—479.

—— 2001, On the paradigms of comparative Uralic studies. — FUF 56, 29—41.

Itämerensuomi — eurooppalainen maa, Jyväskylä 1997 (Studia Historica Fenno-Ugrica II).

J u l k u, K. 2002, Maanjäristys. — Kanava 7, 489—492.

K a l l i o, P., d e S m i t, M. 2002, Missä ovat richterit? — Kanava 9, 634—636.

K o r h o n e n, M. 1996, Typological and Historical Studies in Language. A Memorial Volume Published on the 60th Anniversary of his Birth, Helsinki (MSFOu 223).

K ü n n a p, A. 1997, On the Origin of the Uralic Languages. — Western and Eastern Contact Areas of the Uralic Languages, Tartu (FU 21), 65—68.

—— 1998. Breakthrough in Present-Day Uralistics, Tartu.

—— 2002, Õigeusklike õudusunenägu. — KK, 523—525.

L i g e t i, L. 1986, A magyar nyelv török kapcsolatai a honfoglalás előtt és az Árpád-korban, Budapest.

M a r c a n t o n i o, A. 2000a, The Uralic Language Family. Facts, Myths and Statistics, Oxford—Boston (Publications of the Philological Society 359).

—— 2002b, Comment on the Paper: "On the Paradigms of Comparative Uralic Studies" by Juha Janhunen (FUF 2001, Vol. 56 : 29—41). — FUF 57, 466—470.

M e i l l e t, A. 1967, The Comparative Method in Historical Linguistics, Paris.

N i c h o l s, J. 1992, Linguistic Diversity in Space and Time, Chicago.

P a a s o n e n, H. 1912/1913—1916/1917. Beiträge zur finnischugrisch-samojedischen Lautgeschichte. — Keleti Szemle 13—17, 1—224.

P u s z t a y, J. 1997, Ajatus uralilaisten kansojen ketjumaisesta alkukodista. — Itämerensuomi — eurooppalainen maa, Jyväskylä 1997 (Studia Historica Fenno-Ugrica II), 9—19.

S a m m a l l a h t i, P. 1988, Historical Phonology of the Uralic Languages (with Special Reference to Samoyed, Ugric and Permic). — The Uralic Languages. Description, History and Foreign Influences, Leiden (Handbook of Uralic Studies I).

W i i k, K. 2002, Eurooppalaisten juuret, Jyväskylä.