

Studying the Uralic proto-language

Jaakko Häkkinen (January 27th, 2006; proofread and PDF'ed July 26th, 2025)

[This is my own translation from my Finnish article “Uralilaisen kantakielen tutkiminen”, published in *Tieteessä tapahtuu* 1 / 2006.]

In *Tieteessä tapahtuu* 7 / 2005 Kalevi Wiik presented a fresh study about the genes of English population. At the end of his text Wiik repeated his belief in the theory according to which the very first post-Glacial population in England spoke Uralic proto-language. I would like to clarify a few points about Proto-Uralic and the means of studying it for the readers (nowadays also "Finno-Ugric" often points to the Uralic language family as a whole). I try to keep my presentation as clear and understandable as possible, so that even the readers unaware of the discipline concerned could follow my argumentation.

One of the basic principles of science is that every object is studied by the methods proper and relevant for that particular object. Consequently, language is studied by linguistics, material culture is studied by archaeology, and genes are studied by genetics. Because language, for example, is not connected to any particular gene, we cannot study the language by the methods of genetics: there is no gene that would determine the language we speak.

This is probably not a surprise. And still there are scholars who think they can act differently. Kalevi Wiik, for example, thinks that because we cannot reach the most distant times by the methods of linguistics, we must turn to the methods of other disciplines – such as genetics and archaeology – in order to study the linguistic situation in the distant past (Wiik 2002: 23). Later in this paper we will see that material culture and genes have a few similarities in the mechanism of their heritage, so I bundle them up together under the one and the same method, just like Wiik himself does.

Now the elementary question is: Is it possible to get reliable information about the linguistic past by the method used by Wiik, namely by following the genetic and/or archaeological continuity back in time?

Reliability of the method

The only argument of Wiik goes like this: “The population of Finland descends from the earliest post-Glacial inhabitants. They came from the south, from (Northern) Central Europe. In archaeological data there is perceivable an evident continuity from the earliest inhabitants to the historical era. Thus, the earliest inhabitants of Finland have spoken a Uralic language, predecessor of present-day Finnish. Because those people arrived from Central Europe, the original Proto-Uralic area must evidently have been situated there.”

This sounds logical so far, does it not? Even though the conclusion about language has been made by the means of other disciplines than linguistics. But let us see what other results has been gained with the similar method.

By pleading the archaeological and/or genetic continuity, the original area of Proto-Indo-European has been “proved” to been situated in India, Caucasus, Middle-Asia, Anatolia, Ukraine, and Central Europe (see Mallory 1989: 143–185). Respectively, the same method has been used to “prove” that the original Proto-Uralic area must be located in Siberia (Kosinskaya 2001), the Upper Volga Region (Carpelan 2000), and Central Europe (Wiik 2002).

Naturally all these testimonies cannot be true, because the original area of every proto-language has been narrow (I will return to this later). Not only the place, but also the time concerned is contradictory: Indo-European continuity in Central Europe has been “proved” to reach the Neolithic (Renfrew 1987) and the Palaeolithic Age (Makkay 2001), and Uralic continuity in Finland has been “proved” to reach the Neolithic (Meinander 1984) and the Mesolithic Age (Nuñez 1987).

And above all, the results gained by this method are contradictory also concerning the linguistic identity: The Late Palaeolithic inhabitation of Central Europe has been “proved” both as Indo-European (Makkay 2001) and Uralic (Wiik 2002).

In short: This method (making conclusions about language by means of other disciplines than linguistics) is most unreliable and thus totally worthless. But why is the method so unreliable?

Explanation

First, we must understand that the archaeologically perceivable continuity is evident about everywhere (Mallory 2001: 357) – continuity does not mean that there may not be any external influence, but it means that the external influence is, as archaeologists see it, too weak that it could have caused a language shift.

Genetic continuity is also evident everywhere; the only exception would be an area where the earlier people would have disappeared totally before the arrival of new inhabitants. Only then would there be a clear discontinuity in archaeological and genetic data (if only there were any remains later to compare to).

One archaeological culture can be multirooted, so that influences have flown from the different directions (one item type from here, another from there), and similarly the genetic roots of a person are in theory doubled in every generation (with the exception of paternal and maternal lineage, which I shall discuss later).

Language is, however, a different case: language is always one-rooted. This means that a child adopts one of the languages spoken around him as his mother tongue. This language has always only one root: the root of Finnish leads to Proto-Uralic, and the root of Swedish leads to Proto-Indo-European. Later alien features cannot change the genealogical identity of a language. Even though Finnish has a plenty of shared words and structures with Swedish, it is still a Uralic language (Laakso 1995).

A language is “born” so that at a certain area there occurs enough changes, which differentiate a vernacular from the others spoken by neighbors. It seldom happens that the result is a sharp boundary between two regionally close vernaculars, because people are in contact with each other and might adopt certain features from their neighbors. Birth of a language rather occurs via the disappearance of intermediary dialects. If, for example, dialects 1 and 2 are leveled to unity (they adopt from each other all the features which separated them), the

process results in the growing difference and thus sharpening of the linguistic boundary between dialects 2 and 3. (Salminen 1999: 14; 2001: 385)

It follows that a language is always “born” in a narrow area: The wider the area is, the more improbable is the occurring of a sharp boundary, because the distributions of traits do not match each other as easily as in a narrow area. Those who suggest that the Proto-Uralic speech area has been wide, ignore this linguistic law: Proto-Uralic must have been “born” in a narrow area (Janhunen 1999: 34). And those who suggest that Proto-Uralic was a mixed language born as a result of intensive areal contacts, ignore this very same law: Also mixed languages are born in a narrow area.

The prehistory of Finland is full of influences from different directions at different times. All these have left traces in local cultures; some more, some less. And still the Finnish language has only one root, which leads to Proto-Uralic.

Thus, it follows that when we try to solve with which cultural or genetic wave of influence the Uralic language could be associated, we are merely relying on luck. As we have seen, one scholar thinks the Uralic language has spread to Finland along with the original inhabitants, while the other thinks it is connected with the Neolithic Combed Ware. It is simply impossible to get any reliable information about language by the methods of archaeology or genetics.

Even the one-rooted father and mother lineages cannot help us. There is no single Finnish lineage for these, because all the Finns do not descent from the same paternal or maternal ancestors. The Finnish father lineages point to different directions, and so do the mother lineages. There is no way to find out with which lineage the Uralic language has spread here, so the name of the game here is, again, lottery.

Probability of success

It follows from the one-rootedness of language compared to the multi-rootedness of culture or genome that the wider the area of language, the less the continuity in archaeological or genetic data can actually tell us. Let us suppose that the Uralic language family consists of about 30 speech areas. Proto-Uralic was spoken in one of these areas (unless it was located outside of present-day Uralic area) – to all the other areas Uralic language has spread later. Because Proto-Uralic is a much later language than the end of the Ice Age and it surely did

not spread to empty areas, in all these other 29 presently Uralic areas there must have occurred a language shift: The earlier inhabitants have abandoned their original languages and adopted a Uralic language.

In all these 30 areas the archaeologically perceivable continuity is evident: It has also been used as an argument for locating the original Uralic area, as we saw at the beginning of this article. It follows that archaeological continuity corresponds with linguistic continuity only in one area, when in all the other 29 areas archaeological continuity corresponds with linguistic discontinuity and language shift. So, the probability of success, when trying to locate the original Proto-Uralic area by the results of archaeology is $1/30$, that is 3.33 %.

Therefore, there is 96.67 % chance to fail. No wonder then, that the results gained by this method are contradictory. And in the case of Wiik, who is searching the Proto-Uralic area even outside the known Uralic area, the chance of failing is yet greater.

Solution

What if we nevertheless tried linguistics, even though Wiik believes it cannot tell anything about the times so distant. Perhaps Wiik just is not aware of all the choices available in linguistics. For example, there is no law to determine how far in the past the linguistic method could reach. It depends on the language in question: the areal width of the language family, intensity of the contacts and the width of the contact language family all allow us to follow the language farther in the absolute past. As far as relative past is concerned, the Uralic languages can be traced to the very beginning, that is Proto-Uralic, and for the absolute past this means about 6 000 years (see the references below). [Update in 2025: Nowadays some 4000+ years.]

Because Wiik wants to reach Proto-Uralic and because Proto-Uralic can be reached by the methods of comparative linguistics, it is evident that we take the linguistic results into consideration. After all, it seems reasonable that the linguistic past is best reached by the methods of linguistics – at least the object and the methods would now correspond to each other. We surely would not study the Neolithic physical remains by the linguistic methods any more than we would study the atmosphere of Venus by the methods of dentistry.

For example, linguistics can tell us what kind of language was spoken in Central Europe before the Indo-Europeanization of the area. There have been found features, both phonological and lexical, from the aboriginal languages, which point to the influence of non-Indo-European languages.

Wiik has also presented similar substrate features in Proto-Germanic, Proto-Baltic and Proto-Slavic; he assumes that these features are due to Uralic influence originating from the process when the originally Uralic-speaking inhabitants learned Proto-Indo-European through the filter of their own language system (Wiik 1999). However, such phonetic, prosodic and structural features may well be due to internal development of language; and even if they were substrate features, nothing could prove that they were due to Uralic influence. Furthermore, many of the features presented by Wiik are too late to be Uralic. The hypothesis about the Uralic substrates features in Germanic has been disproved years ago (Kallio 1997b; Kallio, Koivulehto & Parpola 1998).

The lexical evidence is a more reliable indicator of the identity of the substrate language: If those non-Indo-European words would be similar to Uralic words, this would truly be a strong proof for the Uralic identity of those substrate language(s). Especially so, because some words are always very stable and the relation would be perceivable even long after the divergence. For example, Finnish word *kala* ‘fish’ has a cognate in a Samoyed language, Forest Nenets spoken in Siberia: *kal'a* ‘fish’ – although these languages diverged several thousands of years ago.

This is to say that if the aboriginal languages of Central Europe were Uralic-related, these substrate words would be identifiable when compared to the present-day Uralic languages and the Proto-Uralic reconstructed based on their comparison.

It has been found out that in a language-shift situation like the one Wiik supposes particularly the vocabulary concerning local nature and geographical features is exposed to borrowing (Saarikivi 2000). Accidentally, it has become clear that these ancient languages in Central Europe do not resemble Proto-Uralic in the least – neither phonologically nor lexically (Kallio 1997a; 1997b; Schrijver 2001).

Thus, linguistics has proved that Uralic-related languages were not present in Central Europe before Indo-European expansion; the local aboriginal languages were totally distinct from them. The original area of Proto-Uralic was not in Central Europe, nor had it ever even

spread there [before Hungarian much later]. Linguistics can also help us to solve the original Proto-Uralic temporal and spatial location more accurately; very comprehensive and clear guide to the topic is *Suomalaisten esihistoria kielitieteen valossa* by Kaisa Häkkinen (1996). However, this is not a place to sport with that topic – those interested in the question may study it by themselves. It is sufficient to sum up that the linguistic evidence clearly points to eastern origin. Scholars are only arguing if the original area was to the west or east of the Ural Mountains (Salminen 2001: 391; Janhunen 2000: 63). In addition, both Proto-Uralic and Proto-Indo-European seem to be much later languages than Wiik supposes, dated no earlier than in the fourth millennium BC (Kallio 1997a; Carpelan & Parpola 2001).

On the studies for origin

What have we learned about the studies for origin? At least that we cannot get reliable information about the language by any other methods than linguistic. It has also become clear that there is absolutely no support whatsoever for locating Proto-Uralic in Central Europe, not to mention Great Britain.

The method adopted by Wiik and many other scholars – to ignore the best argued linguistic evidence and instead rely on archaeology and/or genetics – has been revealed extremely unreliable. In scientific studies for origin, we must always respect the autonomy of disciplines: If we study material culture, the results of archaeology must form the very basis; if we study language, the results of linguistics must form the basis.

In practice, applied in the Uralic studies, this means that after we have located Proto-Uralic in time and space by linguistic methods, we may take archaeology along. This is done by finding such an archaeological culture which happens to match the proto-language concerned by the time, the place, and the direction of expansion. In short, we will not prefer the lottery method any longer in search of a matching culture. This way Proto-Indo-European has been located in the Pontic steppes at the fourth millennium BC (Carpelan & Parpola 2001 with further references).

If it is not done this way but conclusions about language are made while ignoring the results of linguistics, we are not talking about scientific studies of origin. Then there remains two further options: If the results of linguistics even could not tell anything about the question, it

is just a matter of guessing, the probability of which to hit the correct match is no more than a few percents. If, on the other hand, the results of linguistics could tell a lot about the topic (like in the Proto-Uralic case) but they are still ignored, it is merely a leap outside of science, to the world of fantasy (Saarikivi 2003).

There is no single interdisciplinary method, which could magically solve the problems of linguistic, genetic and cultural origin all by one. Every component of the origin must be studied by methods matching the study object, and only after this can all the independent results be connected as an interdisciplinary summary.

Consequently, the origin of the Finns is not a coherent entity where we could proclaim, after finding the origin of one or two components, that the Origin is now resolved. Such a case would be possible only in the world where genes, culture, and language were always inherited from the one and the same homeland. The populations would have been born as finished packages in their original homeland, and they would not receive any genetic, cultural or linguistic influx during their migration. Populations would all live in the vacuum of their own, inbreeding and lacking contact with other populations. Naturally, this is not the case in our world.

The origin of a people is rather a multileveled and constantly changing puzzle, where the object of study is not identical either in genetic, cultural or linguistic level with the “same” people a thousand years ago. The genetic roots of the Finns lead to many differing directions, and the same goes with cultural roots. Yet our language is quite a late newcomer from the east.

There is no contradiction in such a view, because the components of the origin are not interdependent: They function at totally different levels and thus even cannot contradict with each other. That someone has dark skin does not automatically mean that his mother tongue could not be Finnish. Language, genes, and culture do not actually meet at any level – they meet only in the artificial concept of “people” we use.

This is the very reason why every scholar who understands the origin of certain people as one coherent object of study is automatically misled. There are many origins, and they are totally independent. There is no way to solve the absolute origin of people, because there is no absolute origin at all. By linguistic study we reach only the linguistic origin, by genetic

study only the origin of certain genetic features, and by archaeological study only the origin of certain features of material culture.

Conclusion

In the studies for origin, it has sometimes been pleaded to different schools, as if different views could justify the contradictory results. Whether such schools really exist or not, it remains a fact that some methods are more reliable than others. A school applying an unreliable method is scientifically less worthy than a school applying a more reliable method. An unreliable method will not become any more reliable, no matter how long list of scholars using the method is presented.

Wiik sees the key question to be: “How has occurred such a situation that some of the people linguistically related to Finns are not genetically related to them? How has occurred such a situation that some of the people genetically related to Finns still speak a language not related to Finnish?” (Wiik 2002: 28; my translation.)

Wiik answers, leaning to the method which has no match in unreliability, and which ignores all the plausible results of linguistics, that all those people in Central Europe which are genetically related to Finns have earlier spoken a Uralic language but later shifted it to an Indo-European language.

I, on the other hand, can answer just like anyone else scientifically studying the origin would answer: The first inhabitants of Finland after the last Ice Age arrived mainly from the south, but later they shifted their language to the Uralic one, spreading from the east. Traces of Paleo-European languages earlier spoken in Central Europe have been reached, and those languages were definitely not Uralic.

I believe the reader is now, after this article, able to assess which one of these answers is based on a more reliable method and is thus scientifically more plausible.

[The arguments concerning the unreliability of "continuity"-method are of course relevant in Indo-European studies also; thus we can consider erroneous any urheimat-theory based on archaeological and/or genetic continuity and contradicting the most plausible linguistic

evidence. This includes such recent theories like those of Colin Renfrew (1987), János Mak-kay (2001) and Mario Alinei (https://www.academia.edu/11765622/The_Paleolithic_Continuity_Theory_on_Indo_European_Origins). I am sure the list would finally become very long if someone was patient enough to collect all such theories.

I also recommend for all interested in the subject a critical article by J. P. Mallory concerning the “continuity card” argumentation (Mallory 2001) – it has been a major inspiration for this text.]

Literature

- Carpelan, Christian. 2000. Essay on archaeology and languages in the western end of the Uralic zone. Congressus nonus internationalis Fenno-ugristarum 7.-13.8.2000. Tartto 2000.
- Häkkinen, Kaisa. 1996. *Suomalaisten esihistoria kielitieteen valossa*. Tietolipas 147. Helsinki: Suomalaisen kirjallisuuden seura.
- Janhunen, Juha. 1999. Euraasian alkukodit. In Fogelberg, Paul (ed.), *Pohjan poluilla. Suomalaisen juuret nykytutkimuksen mukaan*. Bidrag till kännedom av Finlands natur och folk, 153.
- Janhunen, Juha. 2000. Reconstructing Pre-Proto-Uralic typology spanning the millennia of linguistic evolution. Congressus Nonus Internationalis Fenno-Ugristarum 7.-13.8.2000 Tartto.
- Kallio, Petri. 1997a. Uralilaisten alkuperä indoeuropeistisesta näkökulmasta. *Virittäjä* 101.
- Kallio, Petri. 1997b. Uralic substrate features in Germanic? SUSA 87, Helsinki 1997.
- Kallio, Petri & Koivulehto, Jorma & Parpola, Asko. 1998. Kantagermaanin suomalais-ugrilainen substraatti: edelleen perusteeton hypoteesi. *Tieteessä tapahtuu* 3 / 1998.
- Kosinskaja, L. L. 2001. The Neolithic period of north-western Siberia: The question of southern connections. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations* (toim. Carpelan et al.). SUST 242.
- Laakso, Johanna. 1995. A spade is always a spade. In Suhonen, Seppo (ed.), *Itämerensuomalainen kulttuurialue*. Castrenianumin toimitteita 49.

- Makkay, János. 2001. The earliest Proto-Indo-European-Proto-Uralic contacts: An Upper Palaeolithic model. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. SUST 242.
- Mallory, J. P. 1989. *In Search of the Indo-Europeans. Language, Archaeology and Myth*. London: Thames and Hudson.
- Mallory, J. P. 2001. Uralics and Indo-Europeans: Problems of time and space. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. SUST 242.
- Meinander, C. F. (ed.). 1984. *Kivikautemme väestöhistoria*. In Gallén, Jarl (ed.), *Suomen väestön esihistorialliset juuret: Tvärminnen symposium 17.–19.1.1980*. Bidrag till kännedom av Finlands natur och folk, 131.
- Nuñez, Milton G. 1987. A Model for the Early Settlement of Finland. *Fennoscandia Archaeologica* 4.
- Parpola, Asko. 1999. Varhaisten indoeurooppalaiskontaktien ajoitus ja paikannus kielellisen ja arkeologisen aineiston perusteella. In Fogelberg, Paul (ed.), *Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan*. Bidrag till kännedom av Finlands natur och folk, 153.
- Renfrew, Colin. 1987. *Archaeology and Language. The Puzzle of Indo-European Origins*. Harmondsworth: Penguin Books Ltd.
- Saarikivi, Janne. 2000. Kontaktilähtöinen kielenmuutos, substraatti ja substraattinimistö. *Virittäjä* 104.
- Saarikivi, Janne. 2003. Fiktiivistä tiedettä? *Hiidenkivi* 1 / 2003.
- Salminen, Tapani. 1999. Euroopan kielet muinoin ja nykyisin. In Fogelberg, Paul (ed.), *Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan*. Bidrag till kännedom av Finlands natur och folk, 153. Helsinki 1999.
- Schrijver, Peter. 2001. Lost languages in northern Europe. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. SUST 242.

Wiik, Kalevi. 1999. Pohjois-Euroopan indoeurooppalaisten kielten suomalais-ugrilainen substraatti. In Fogelberg, Paul (ed.), *Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan*. Bidrag till kännedom av Finlands natur och folk, 153.

Wiik, Kalevi. 2002. *Eurooppalaisten juuret*. Jyväskylä: Atena.

SUSA = Suomalais-Ugrilaisen Seuran aikakauskirja / Journal de la Société Finno-Ougrienne.

SUST = Suomalais-Ugrilaisen Seuran toimituksia / Mémoires de la Société Finno-Ougrienne.